The Dissertation Committee for Darren Christian Gay Certifies that this is the approved version of the following dissertation:

Inside the microbial weapons factory: Structural studies of polyketide biosynthetic machinery

Committee:

Adrian Keatinge-Clay, Supervisor

Jon Robertus

Dave Hoffman

Jeffrey Barrick

John Ekerdt

Inside the microbial weapons factory: Structural studies of polyketide biosynthetic machinery

by

Darren Christian Gay, B. S. Bioch.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2015

Dedication

I would like to dedicate this work to my family.

To my beautiful wife Caitlyn, for your unending support and love. You mean more to me than you can ever imagine, and I will love you forever.

To the Stanzeski's: Chester, Dorothy, Chester Jr. and Jan. You have all been so wonderful to me, I couldn't have asked for a more loving family to marry into. Thank you so much for your support and encouragement during my graduate career.

To Dad, who taught me some of the most essential things in life, like how to tell a story, bait a crawfish basket, and order a steak. I know I was supposed to be a fighter pilot, but hey, there's still time.

To Glen, who has saved my life more than once and whose labor contributed to a generous fraction of the following dissertation. Your advice is priceless, and I have learned more from you than I'm sure you're aware.

And to Mom, whose constant barrage of support and encouragement pretty much makes it impossible to fail at anything. I know you must be proud as you hold this dissertation, but what else would you expect after telling someone for 31 years that they can do anything they set their mind to? Every shred of confidence, self-reliance, cleverness, and hope that I have, I have learned from you.

Τίς εἶναι θέλεις, σαυτῷ πρῶτον εἰπέ: εἶθ' οὕτως ποίει ἃ ποιεῖς.

-Ἐπίκτητος

Acknowledgements

I would like to thank all of my lab mates who worked with me at the bench, from whom I've collected so much experimental insight. Thanks to Amanda Hughes, my very first mentor in practical biochemistry (I still don't know how you got stuff to work). Thanks to Shawn Piasecki, who kept the lab light-hearted and all of us well-fed. Thanks to Andy Harper, and to your amazing talent for recalling advanced biochemical knowledge that most of us would normally have to look up. Thanks to Eta Isiorho-Mansoorabadi, for your colorful character, sharp wit, and clever suggestions to get around problems. Thanks to Chris Fage, with whom I've had the pleasure to ride tricycles around a synchrotron. Thanks to Constance Bailey, who always had a great laboratory tip or trick to share with the rest of the group. Thanks to Drew Wagner, a great scientist to collaborate with and an outstanding wrestling opponent. Thanks to Jessica Meinke, your ability to learn and apply experimental techniques is excellent, and I see great potential for you as a professional scientist. Thanks to Drs. Jianting Zheng, Jia Zeng, Abram Axelrod and Cole Stevens, a set of amazingly talented post-docs whose time in the Keatinge-Clay lab overlapped with mine and from whom I have learned very much. I would also like to thank Phillip Spear, who volunteered time and energy as an undergraduate research assistant and for his contributions to the isomerase project. Nathan Crandall has also been an incredibly helpful undergraduate research assistant, and his scientific curiosity will undoubtedly produce a stellar physician.

I would also like to thank the many scientific mentors who have contributed to my training and education. In 2001, Drs. John Mike White, Yangming Sun, Qi Wang, and John Ekerdt took me in as an undergraduate research assistant and constructed the foundation of how I would approach science for the rest of my life. It was in your labs

that I first fell in love with science, when Qi answered my question of "Why are we doing this experiment?" with the answer "Because no one has ever done this before, and no one knows what will happen when we do it." I doubt I would have ever pursued a PhD without the experience I had as an apprentice in your labs, and I will forever be thankful. On that note, I would never have *completed* a PhD without the help and valuable input from my dissertation committee, Drs. Dave Hoffman, Jon Robertus, Jeffrey Barrick, and John Ekerdt. Thank you for all the helpful feedback during committee meetings and in regards to the following dissertation, I feel exceptionally honored to have each of your signatures on this document.

Finally, I would like to thank my advisor, Adrian Keatinge-Clay. It's incredible to think that just six years ago I introduced myself to you as a curious undergraduate, and today we have published a number of scientific papers together and have very exciting plans on our horizon. The opportunity you have given me in life is invaluable, and I will never be able to adequately express my gratitude for the physical, chemical, and biological training I received under your guidance. And the training I received from you did not stop at the sciences; you are truly the man who Robert Heinlein alludes to when he mentioned that "specialization is for insects". I have always been impressed by your ability to simultaneously manage a lab of blossoming scientists, teach advanced courses in molecular biophysics, convince the government to fund our lab with research grants, write, edit, and publish multiple manuscripts, organize kickball tournaments and lab outings, all while juggling the behind-the-scenes responsibilities that come with being a research professor that many of us never get to appreciate. Thank you for the opportunity of a lifetime, I am eternally grateful.

Inside the microbial weapons factory: Structural studies of polyketide biosynthetic machinery

Darren Christian Gay, PhD

The University of Texas at Austin, 2015

Supervisor: Adrian Keatinge-Clay

Polyketides are a class of small molecules synthesized by a broad spectrum of bacteria, plants, and fungi, and many exhibit powerful bioactive properties. The number of clinically-relevant compounds adapted from polyketide scaffolds is growing, eliciting attempts from synthetic organic chemists to construct polyketide-related compounds in the laboratory from simple chemical building blocks. Unfortunately, the current efficiency by which a skilled artisan can synthesize even small quantities of a polyketide is severely limited by the functional and stereochemical complexity of these compounds. Conceptually, it would be much simpler to genetically reprogram the enzymes responsible for polyketide biosynthesis to produce designer molecules; however, the massive size of polyketide synthase enzymes has hindered efforts towards understanding critical features of their structures and mechanisms. Only very recently has structural information become available for enzymes involved in polyketide biosynthesis, providing an initial glimpse into the inner workings of these subcellular pharmaceutical factories. It will not be possible for mankind to fully realize the potential of engineered polyketide synthases without understanding how their architectures govern the molecules they have evolved to produce.

In this work, the structure and mechanism of several enzymes involved in polyketide biosynthesis is investigated. An unprecedented architecture for the ketoreductase-enoylreductase didomain from the second module of the spinosyn polyketide synthase reveals structural divergence from the related mammalian fatty acid synthase, and reconstituted *in vitro* activity of the enoylreductase domain indicates the isolated enzyme retains activity apart from its parent polyketide synthase module. The dehydratase domain isolated from the tenth module of the rifamycin polyketide synthase, previously hypothesized to only form double bonds with (*Z*) geometry, was found to have altered stereoselectivity dependent on the carrier handle bound to the substrate. The enoyl-isomerase domain, isolated from the fourteenth module of the bacillaene polyketide synthase, utilizes a catalytic mechanism that relies only on a single active site histidine. A series of ketosynthase domains from trans-acyltransferase polyketide synthases reveal how polyketides bind covalently to the active site of the ketosynthase, and how the flanking subdomain of the ketosynthase is used as an anchor point for the assembly of the polyketide synthase megacomplex.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

During the fall of 1928, the Scottish scientist Alexander Fleming noticed something strange in one of his petri dishes. A blue-green mold had occupied a region of the culture plate that should have been inhabited by a species of *Staphylococcus*. Upon closer inspection, he noticed that the growth of *Staphylococcus* was inhibited when in close proximity to the mold, indicating that the mold was producing a substance that prevented the bacteria from encroaching upon its space. If the mold was producing a compound that is toxic to *Staphylococcus* yet harmless to humans, perhaps this substance could be isolated and used as an antibiotic? Unknowingly, Dr. Fleming had made one of the most critical findings responsible for ushering in the modern era of natural product discovery; penicillin was scheduled for mass production in the United States just 12 years later. It was not until 1945 that scientists began to understand penicillin's mechanism of action, when Dorothy Hodgkin used X-ray crystallography to determine the atomic structure of this small molecule. It is astounding that just within the last one-hundred years man has come to the realization that there is an arsenal of molecular weapons under our feet that have been rigorously developed by natural selection for billions of years. We have adopted some of these weapons for our own uses, to augment our immune systems against infection, lower cholesterol, and even fight cancer. We have discovered that X-ray crystallography allows us to visualize the location of every atom in these natural products, permitting scientific inquiry of the mechanism of action, and form hypotheses concerning potential chemical modifications to increase efficacy. We are currently embarking on a new frontier of reengineering the biological machinery that has evolved to construct these molecules, as discovering natural products gives way to designing natural products.

Polyketides are a class of natural products that have shown tremendous bioactive potential (**Figure 1-1**). For example, the Buruli ulcer afflicts thousands in developing countries with painful skin lesions, resulting from an infection by *Mycobacterium ulcerans* (Nakanaga *et al.*, 2013). This pathogenic mycobacterium produces the polyketide mycolactone, inhibiting the host's immune response and causing severe dermal ulcers. Ironically, a common treatment for the Buruli ulcer is rifamycin, a polyketide antibiotic produced by the soil-dwelling bacterium *Amycolatopsis mediterranei*, also used as a first line of defense against tuberculosis (World Health Organization, 2010). Perhaps the most comprehensively studied polyketide is the antibiotic erythromycin, generally prescribed for bacterial infections for those who have an allergy to penicillin (**Figure 1-2**) (Khosla *et al.*, 2007). Although a powerful antibiotic, erythromycin also causes abdominal disturbances in some patients. Conceivably, one of the functional groups of erythromycin could be altered, and the new analog may retain antibiotic properties without adverse side effects. This hypothesis would be very challenging to test; erythromycin is a 14-membered macrolide containing 10 chiral centers and decorated with two sugar moieties. The synthesis of erythromycin was eventually completed by the Nobel laureate Robert Woodward, a heroic effort that was not accomplished until thirty years after its discovery (Woodward *et al.*, 1981). If organic chemists were to develop a modification to erythromycin that revealed enhanced medicinal properties in humans, it would not be economically possible to synthetically generate the kilograms of analog required for distribution as a prescription antibiotic. A solution to this problem potentially lies in genetic engineering: rewiring and reprogramming the polyketide biosynthetic machinery from the host organism to produce natural product analogs that can be isolated by fermentation and extraction. If scientists can harness and manipulate polyketide synthases (PKSs) at the genetic level, the number

of polyketide compounds that would become accessible for research and production would be virtually unlimited.

Type I PKSs are multienzyme polypeptide complexes that represent some of the largest proteins discovered, and provide an exceptional opportunity to examine enzyme catalysis, molecular recognition and protein-protein docking (Shen, 2003; Keatinge-Clay, 2012). The enzymatic domains employed by PKSs are similar to those found in fatty acid biosynthesis, and the set of domains required for a single round of polyketide extension and processing is called a module (**Figure 1-3**). The domains present in a module with a complete set of processing enzymes include the ketosynthase (KS) that catalyzes a thio-Claisen condensation of the polyketide intermediate with an α-carboxylated extender unit, the acyltransferase (AT) that selects the α-carboxylated extender unit from the cellular pool of available acyl-CoAs, the ketoreductase (KR) that stereospecifically reduces the newly incorporated β-carbonyl and sets the chirality of the α-substituent, the dehydratase (DH) that eliminates an α-proton and β-hydroxyl as water, the enoylreductase (ER) that stereospecifically reduces the resulting enone to generate a fully saturated intermediate, and the acyl carrier protein (ACP) that shuttles the growing polyketide between all the domains within its cognate module, as well as transfer to the subsequent module. A thioesterase (TE) appended to the terminal module of the synthase releases the polyketide via thioester hydrolysis or macrocyclization. The mammalian fatty acid synthase (FAS) can be thought of as an iterative single-module polyketide synthase, equipped with a complete set of processing domains (Maier *et al.*, 2008). Type I PKSs diverged into two distinct classes early in evolution: *cis*-AT and *trans*-AT (Piel, 2010). The *cis*-AT class was discovered earlier and is better studied than the *trans*-AT class, and is so named because this class harbors an AT domain integrated into the PKS polypeptide C-terminal to the KS, yielding KS-AT-DH-ER-KR-ACP as the domain

architecture for a complete module. *Trans*-AT PKSs do not include an AT domain integral to the module, and AT activity is supplied by a discretely-encoded AT domain that docks to the synthase. *Trans*-AT PKSs have also been observed to form membrane-bound megacomplexes, in which many copies of the megasynthase aggregate to form an organelle-sized structure with the purpose of producing and exporting polyketides (Straight *et al.*, 2007).

PKSs are able to generate such an incredibly diverse array of natural product metabolites utilizing only these core enzymatic domains by taking advantage of two key features: domain deletions and modularity. If a module lacks one of the processing domains (or if it is a null mutant), the α- and β-positions of the polyketide substrate cannot be reduced to a completely saturated acyl intermediate. For example, if a module has the architecture KS-AT-DH-KR-ACP (lacking an ER domain), the resulting polyketide will be transferred to the subsequent module as the enone intermediate. In a similar manner, if the modular architecture was KS-AT-KR-ACP, the polyketide will be transferred as a β-hydroxyl intermediate. The advantage of PKS modularity is related to the concept of assembly line construction: the intermediate generated by module 1 is always passed to module 2 and then to module 3, until the polyketide encounters a module harboring a TE and is released from the synthase. This allows for the relatively rapid reorganization of a PKS by genetic recombination, producing an entirely distinct polyketide from the same set of core enzymatic domains.

With such an elegant yet simple design for natural product evolution, it would seem that man would have the capacity to construct any polyketide imaginable by building a custom PKS through genetic engineering. Instead of the daunting task of synthetically reversing one of the ten asymmetric centers of erythromycin, the KR responsible for processing that stereocenter could presumably be replaced with a KR that

generates the opposite stereochemistry, and the new analog collected by fermentation. Although this endeavor has been successful in some cases, attempts to modify PKSs by domain- or module-swapping generally only produce nonfunctional synthases (Kao *et al.*, 1998; Tsuji *et al.*, 2001; Kumar *et al.*, 2003). Several problems may contribute to this observation: downstream modules may refuse to accept the altered substituent of the polyketide intermediate, or the protein-protein interactions between modules (or within modules) may be disrupted and lead to misfolded domains. The crystal structure of a complete module would undoubtedly solve many problems associated with rational PKS engineering, but no such structure has been published to date. Instead, the domains that compose the PKS module have been isolated and structurally investigated to provide a mosaic body of information about the synthase. The following chapters describes several of these studies, in which domains or didomains dissected from PKS modules are isolated, crystallized, and biochemically interrogated to determine their contribution to PKS enzymology.

In **Chapter 2**, the crystal structure and reconstituted *in vitro* activity of a KR+ER didomain from the second module of the spinosyn PKS (Spn(KR+ER)2) is presented. Despite the similarities in domain architecture between Type I PKSs and the FAS, the structure reveals that the interface shared by the KR and ER domains in Spn(KR+ER)2 is remarkably different than observed in the FAS. The structure indicates ER domains of Type I PKSs are not dimeric, unlike the closely related FAS. It is hypothesized that this architecture is necessary to accommodate the assembly line nature of the megasynthase, permitting the ACP from one module to access the KS of the subsequent module. To confirm that the crystal structure of Spn(KR+ER)2 is representative of the native orientation of the didomain in solution, small angle X-ray scattering (SAXS) analysis was performed. The low-resolution structural envelope generated from SAXS fits the crystal

structure of Spn(KR+ER)2, but does not fit the isolated KR+ER didomain of the FAS, indicating the crystal structure of Spn(KR+ER)2 is not a crystallographic artifact. The active site residues of the ER domain directly responsible for catalysis were interrogated by designing an *in vitro* activity assay coupled with site-directed mutagenesis. It was found that an active site lysine contributes most to catalysis, although the lysine to alanine mutant was able to reduce enoyl substrate mimics to a lesser degree.

In **Chapter 3**, the crystal structure and reconstituted *in vitro* activity of a DH domain from the tenth module of the rifamycin PKS (RifDH10) is presented. Double bonds are installed by PKS modules lacking an ER domain, and are frequently encountered polyketide functional groups. The vast majority of polyketide double bonds have (*E*) geometry, but the macrolactam polyketide antibiotic rifamycin contains a double bond with (*Z*) geometry. The atypical geometry of this double bond was hypothesized to result from a *cis*-dehydration event catalyzed by RifDH10, as this DH domain resides in the module responsible for tailoring the propionyl-derived extender unit that harbors the (*Z*) double bond in rifamycin. To better understand polyketide (*Z*) double bond formation, RifDH10 was isolated, purified, and crystallized. The active site of RifDH10 is structurally homologous to other PKS DH domains that have been solved, and the catalytic dyad (composed of histidine and aspartate residues) is positioned identically to what has been observed in DHs that generate (*E*) double bonds. Despite the structural homology of the RifDH10 active site to *trans*-dehydrating DHs, *in vitro* activity assays were conducted with substrate mimics to further establish the geometry of the dehydration product. RifDH10 is specific for substrate mimics harboring a (3*R*)-hydroxyl if the thioester handle is *N*-acetyl-cysteamine, pantetheine, or a non-cognate ACP (EryACP6). In contrast, RifDH10 is specific for (3*S*)-hydroxyl substrates if the thioester handle is the cognate ACP, RifACP10. Regardless of the thioester handle or the substrate

assayed, RifDH10 only catalyzed the formation of double bond products with (*E*) geometry, indicating the (*Z*) double bond characteristic of rifamycin is likely formed during the subsequent macrolactamization step catalyzed by RifF.

In **Chapter 4**, a modification to the popular expression vector pET-28 is presented, describing how the multiple-cloning region of this vector was replaced with a ligation independent cloning cassette. pET-28 is a commonly employed expression vector for the overproduction of protein in *E. coli*, however the requisite use of restriction enzymes to introduce genes of interest increases costs and reduces yield relative to ligation independent cloning techniques. Homologous recombination in *Saccharomyces cerevisiae* was employed to exchange the multiple-cloning region of pET-28 with a customized ligation independent cloning cassette, generating pGAY-28. The cloning efficiency of this vector is much higher than pET-28, and the protocol used to introduce a gene of interest does not require restriction enzymes. To demonstrate the increased efficiency, a 10.3 kb insert (coding for a complete PKS polypeptide) was successfully incorporated into the ligation independent cloning cassette of pGAY-28, while methods employing restriction enzymes to incorporate an identical insert into pET-28 were unsuccessful. The strategy employed to accomplish plasmid modification via homologous recombination is uniquely versatile, and can be utilized to reconstruct large or small regions of expression vectors easily and quickly.

In **Chapter 5**, the crystal structure and reconstituted *in vitro* activity of an enoyl-isomerase domain (EI) from the fourteenth module of the bacillaene PKS (PksEI14) is presented. EIs are relatively rare PKS processing domains, currently only discovered in four *trans*-AT PKS pathways. They catalyze the reversible isomerization of double bonds from the classical α,β-position to the β,γ-position (relative to the thioester handle of the intermediate). PksEI14 represents the first crystal structure of a PKS EI domain,

revealing a histidine and tyrosine in the active site cavity. To determine the mechanism of isomerization, substrate mimics were synthesized and enzymatic assays were carried out in either H_2O or D_2O. A tyrosine mutant reveals that the active site histidine catalyzes isomerization unaided, and assays conducted in deuterated solvent show that water is not activated by the enzyme as an acid-base catalyst. PksEI14 represents the first example of an enzyme employing a single histidine to independently shuttle a substrate proton to catalyze $\alpha,\beta \rightarrow \beta,\gamma$ double bond isomerization.

In **Chapter 6**, the crystal structure of a KS from the second module of the bacillaene PKS (PksKS2) is presented. KSs from *trans*-AT PKS pathways act as gatekeepers, being specific for the substituents and geometry of the α- and β-positions of the polyketide intermediate. Understanding how this specificity is conferred is imperative if *trans*-AT KSs are to be engineered into unnatural biosynthetic pathways for the production of novel metabolites. The 1.95 Å crystal structure of PksKS2 provides the highest resolution available for a Type I PKS KS, and a cysteine-to-serine active site mutant acylated by the natural polyketide intermediate for PksKS2 reveals key residues that confer substrate specificity. A hexanoyl group was also crystallographically captured in the active site, revealing that both the natural substrate and a tolerable substrate mimic bind in the same orientation. The structure also reveals that *trans*-AT KSs harbor a flanking subdomain homologous to the ordered linker that stabilizes the integrated AT domain of *cis*-AT PKSs. It has been previously hypothesized that this region likely forms the docking site for a *trans*-AT, so that the architecture of a complete module for both the *cis*-AT and *trans*-AT PKS systems would be structurally similar. However, *in silico* docking of a *trans*-AT with the flanking subdomain of PksKS2 indicates significant rearrangement of structural motifs within the flanking subdomain would be necessary to

accommodate the KS:AT orientation observed in *cis*-AT systems. The role of the *trans*-AT KS flanking subdomain is further investigated in Chapter 7.

In **Chapter 7**, seven crystal structures of *trans*-AT PKS KS domains are presented, and their role in stabilizing the *trans*-AT PKS megacomplex is analyzed. The flanking subdomain for each of the seven structures harbors a three-helix motif that self-associates in each of the available crystalline states. Analysis of several *trans*-AT KS structures previously deposited into the Protein Data Bank (PDB) reveals that this three-helix motif also self-associates, forming crystal contacts that stabilize one dimension of the protein crystal. In total, ten distinct crystal structures exhibiting the same well-defined interaction indicate that this crystal contact is likely not an arbitrary artifact of crystallization, but instead serves the role of stabilizing the *trans*-AT PKS megacomplex observed by others with cryo-EM. Even though the amino acid sequences of the three-helix LINKS (**L**ocalization **IN**ducing **K**etosynthase **S**equence) motifs are not conserved in multiple sequence alignments, each crystallographic example demonstrates that LINKS bind to spatially-reversed copies of themselves through ionic and hydrophobic interactions. The lack of LINKS sequence conservation imparts a code that promotes the assembly of an enzymatic network based on homotypic KS:KS lateral interactions.

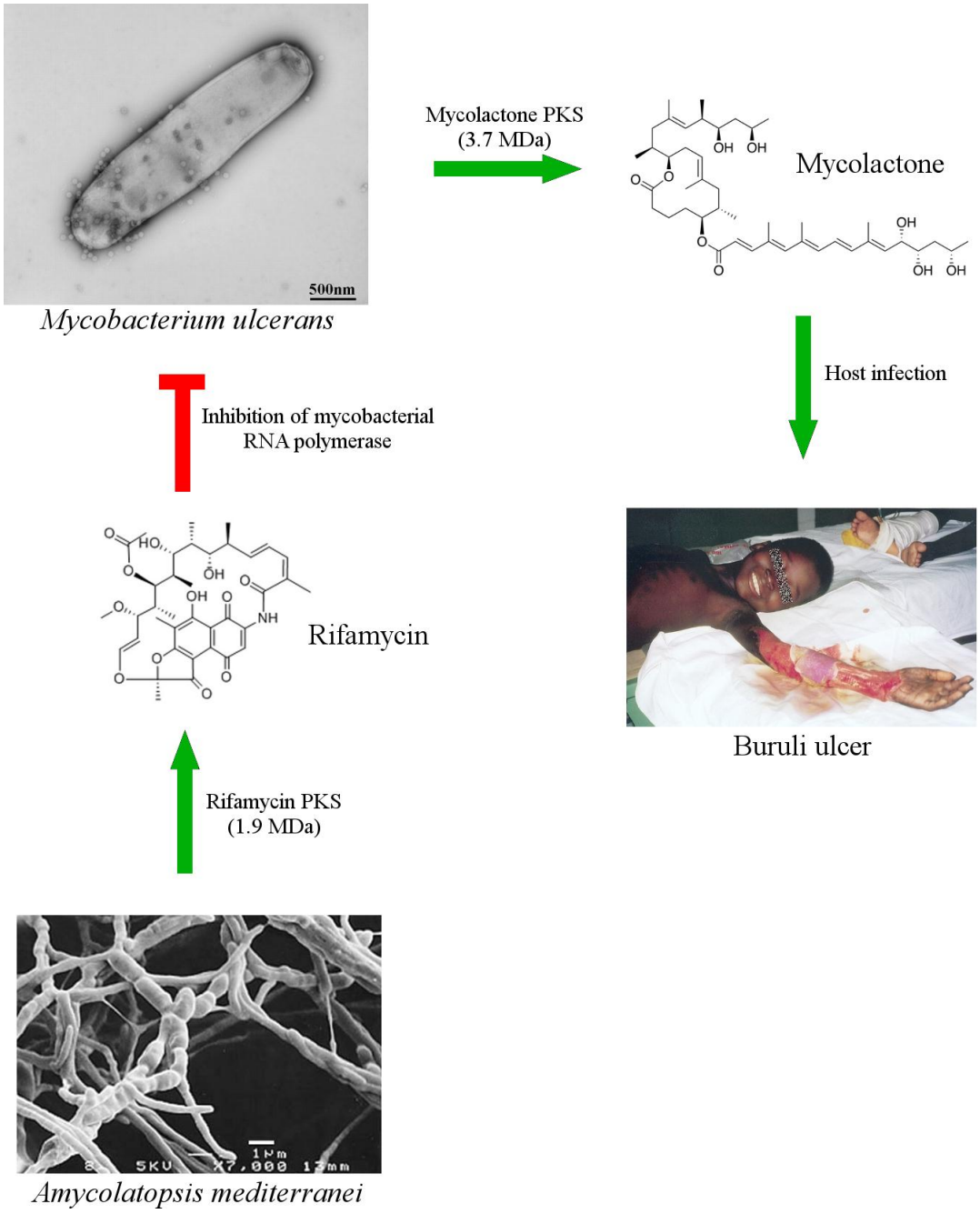


Figure 1-1. Polyketide warfare. *Mycobacterium ulcerans* contains a large plasmid harboring the 110-kilobase gene cluster responsible for the biosynthesis of the toxic polyketide mycolactone. Infection results in severe skin lesions and dermal necrosis. The bacterium *Amycolatopsis mediterranei* produces the polyketide rifamycin, used in the treatment of both the Buruli ulcer and tuberculosis due to its ability to inhibit DNA-dependent RNA synthesis in various mycobacteria.

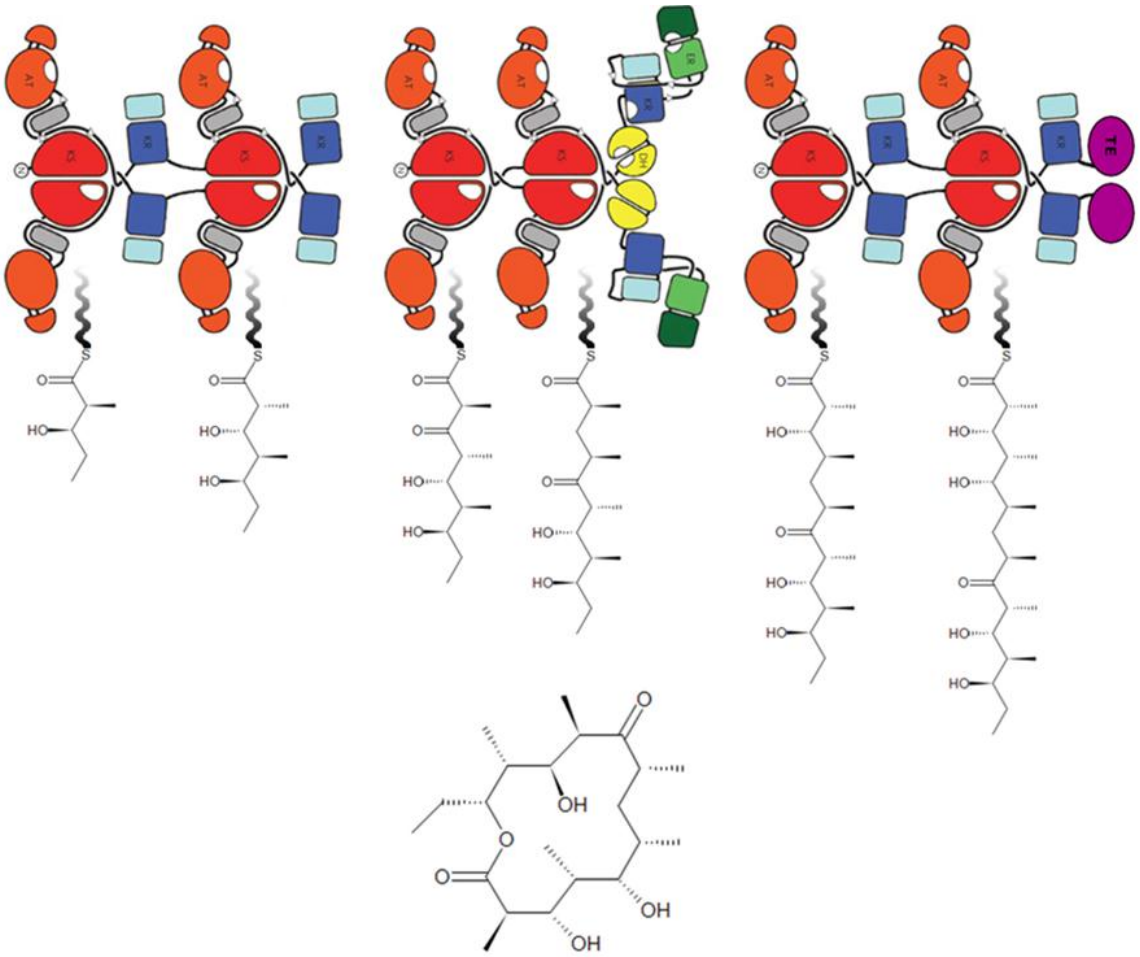


Figure 1-2. The erythromycin PKS. In an assembly-line fashion, 6-deoxyerythronolide B (the aglycone of erythromycin) is constructed stepwise by six PKS modules on three polypeptides. The black line that connects the polyketide to the PKS represents the ~18 Å phosphopantetheinyl arm of the ACP, responsible for shuttling the growing polyketide to all of the domains within the module, as well as transfer to the subsequent module. A TE domain appended to the end of the synthase releases 6-deoxyerythronolide B by lactonization, forming the macrocyclic core of this potent antibiotic.

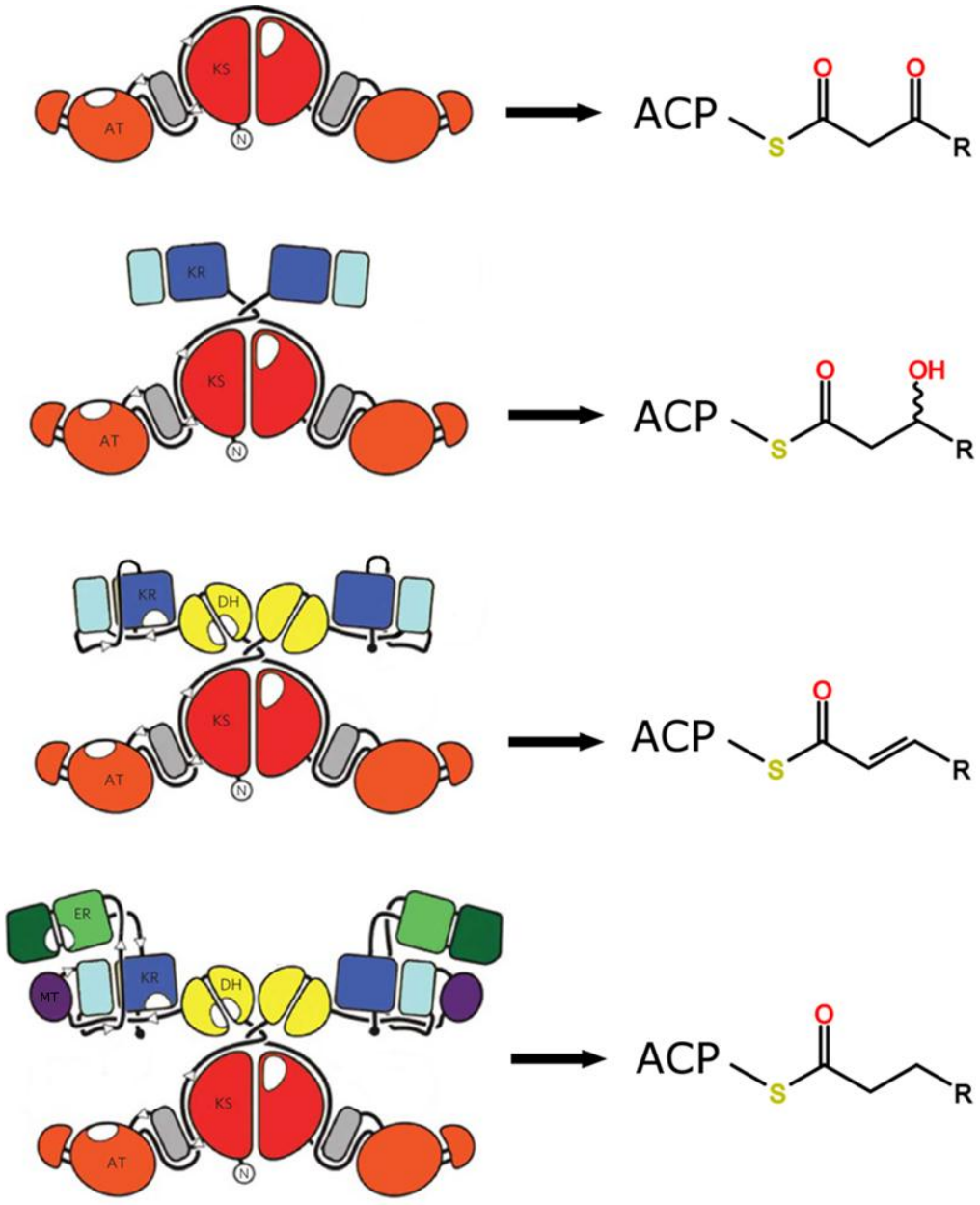


Figure 1-3. Polyketide functional groups depend on module type. The four general classes of PKS modules: α-modules have domain architecture KS+AT+ACP, and generate β-keto products, β-modules include a KR domain and generate β-hydroxy products, γ-modules include a DH domain and generate α,β-olefin products, and δ-modules include an ER domain and generate completely saturated products (Keatinge-Clay, 2012).

Chapter 2. Divergence of multimodular polyketide synthases revealed by a didomain structure

ABSTRACT

The enoylreductase (ER) is the final common enzyme from modular polyketide synthases (PKSs) to be structurally characterized. The 3.0 Å–resolution structure of the didomain comprising the ketoreductase (KR) and ER from the second module of the spinosyn PKS reveals that ER shares an ~600-Å^2 interface with KR distinct from that of the related mammalian fatty acid synthase (FAS). In contrast to the ER domains of the mammalian FAS, the ER domains of the second module of the spinosyn PKS do not make contact across the two-fold axis of the synthase. This monomeric organization may have been necessary in the evolution of multimodular PKSs to enable acyl carrier proteins to access each of their cognate enzymes. The isolated ER domain showed activity toward a substrate analog, enabling us to determine the contributions of its active site residues.

INTRODUCTION

Modular PKSs are megasynthases that synthesize biologically active polyketides such as the antibacterial erythromycin, the antifungal amphotericin and the insecticide spinosyn (Khosla *et al.*, 2007; Smith & Tsai, 2007). Each module along the PKS assembly line harbors a set of enzymes responsible for extending the polyketide chain and setting the chemistries and configurations of the newly added α- and β-carbons.

Chapter 2 was adapted from the following published work: Zheng J, Gay DC, Demeler B, White MA, and Keatinge-Clay AT. (2012) Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615-621. Each author contributed to the experimental design and manuscript preparation.

Although studying enzymes that naturally operate within a megasynthase is inherently challenging, these component enzymes must be more thoroughly characterized to decipher the mechanisms by which polyketides are synthesized. Our efforts have focused on the processing enzymes: a KR that stereoselectively reduces the β-keto group, a dehydratase (DH) that performs an elimination reaction to generate a double bond and an ER that stereoselectively reduces that double bond (Keatinge-Clay, 2007; Keatinge-Clay, 2008; Zheng & Keatinge-Clay, 2011; Zheng *et al.*, 2010; Keatinge-Clay & Stroud, 2006).

Relatively little structural and functional information exists for ER, even though the atomic-resolution structures of the other common PKS domains have been determined and *in vitro* studies have been conducted with those isolated domains (Keatinge-Clay, 2008; Keatinge-Clay & Stroud, 2006; Tsai *et al.*, 2001; Tang *et al.*, 2006; Alekseyev *et al.*, 2007). The crystal structure of the nucleotide-binding subdomain of the ER from the phthiocerol polyketide synthase (PpsC) reveals little about the active site or the overall architecture of ER11. Experiments to identify the key ER residues through mutagenesis of engineered triketide lactone synthases have been performed, and, though residues mediating stereoselectivity were found, a point mutation that abolished activity was not identified (Kwan *et al.*, 2008; Kwan & Leadlay, 2010).

The processing enzymes from the second module of the spinosyn PKS—SpnKR2, SpnDH2 and SpnER2 (named according to their PKS and module of origin)—sequentially process a triketide intermediate during the synthesis of spinosyn (**Figure 2-1**) (Kirst, 2010). Although the SpnER2 domain was not amenable to crystallization, we hypothesized that the Spn(KR+ER)2 didomain might be, as the ER domain inserts into a

loop connecting the KR subdomains and shares an interface with KR within the mammalian FAS (Maier *et al.*, 2008). Indeed, crystals were obtained that diffracted to 3.0 Å. The structure shows that SpnKR2 and SpnER2 both fold similarly to their mammalian FAS counterparts (Maier *et al.*, 2008; Oppermann *et al.*, 2003; Persson *et al.*, 2008). To our surprise, however, the domains form a completely different interface. The conformation adopted by helix αF from SpnER2 precludes the dimerization mode observed for the ER of the mammalian FAS. The three-dimensional organization of the KR and ER domains, confirmed by small-angle X-ray scattering (SAXS), reveals that the complete modules of multimodular PKSs have a different architecture than the mammalian FAS. The first *in vitro* biochemical studies of an ER domain isolated from its PKS show that SpnER2 retains activity on the substrate analog crotonyl-pantetheine and that a lysine contributes most to catalysis.

RESULTS

The overall architecture of Spn(KR+ER)2

The structure of the Spn(KR+ER)2 didomain (SpnB residues 1216–1989; **Figure 2-1a**) was solved to 3.0-Å resolution through selenomethionine incorporation and multiwavelength anomalous dispersion phasing (**Figure 2-2** and **Table 2-1**). The two Spn(KR+ER)2 didomains in the asymmetric unit adopt the same conformation with 0.6-Å Cα r.m.s. deviation (**Figure 2-3a**). Although the SpnKR2 domains have similar average B-factors (49 $Å^2$ versus 56 $Å^2$), the average B-factor of one SpnER2 domain is highly elevated compared to the other (96 $Å^2$ versus 62 $Å^2$; **Figure 2-3b**). The 60 C-

terminal residues of each Spn(KR+ER)2 molecule in the asymmetric unit fold onto the other molecule (that is, it is domain swapped), resulting in the close association of SpnKR2 catalytic subdomains and the occlusion of the SpnKR2 NADPH-binding sites (**Figure 2-4**). Nonphysiological domain swapping is sometimes observed in crystals of truncated proteins (Liu & Eisenberg, 2002; Kishan *et al.*, 1997). Aside from the four residues spanning the domains, the swapped C-terminal residues are folded as expected from related KR structures and have B-factors equivalent to the surrounding protein (Keatinge-Clay, 2007; Zheng & Keatinge-Clay, 2011; Zheng *et al.*, 2010; Keatinge-Clay & Stroud, 2006).

The interface between SpnER2 and SpnKR2 is very different from and ~50% larger (~600 $Å^2$) than the KR-ER interface of the mammalian FAS (**Figures 2-2a**,**b**, **2-5** and **2-6**) (Kirst, 2010). Contacts between the catalytic subdomain of SpnKR2 and the substrate-binding subdomain of SpnER2 form the majority of the interface, which is primarily composed of nonconserved, hydrophilic residues (**Figures 2-5**, **2-6**, and **2-7**). SpnER2 inserts into a loop between the catalytic and structural subdomains of SpnKR2; the N- and C-terminal ends of the insertion loop, referred to here as the N- and C-terminal linkers, help mediate the interface between SpnKR2 and SpnER2. The N-terminal linker (residues 176–189, numbering based on first residue visible in the electron density maps) spans from the C-terminal end of the SpnKR2 structural subdomain to β-strand 1 (β1) of the SpnER2 substrate-binding domain via unconserved residues; the C-terminal linker (residues 503–510) connects β9 of the SpnER2 substrate-binding domain to the N-terminal end of the SpnKR2 catalytic subdomain via semicon-

served residues. Although the N-terminal linker is comparable in length to that of the mammalian FAS (~13 residues), the C-terminal linker is much shorter than its mammalian FAS counterpart (8 versus ~29 residues) and is constrained by a conserved hydrophobic interaction with the KR catalytic subdomain (mediated by Trp506 in Spn(KR+ER)2; **Figures 2-2a**,**b**, **2-6**, **2-7**, and **2-8)** (Maier *et al.*, 2008). Within the mammalian FAS, the separation between the C-terminal end of β9 of ER (residue 1855) and the N-terminal end of the KR catalytic subdomain (residue 1886) is 36 Å. The short C-terminal linker of multimodular PKSs cannot span this distance, preventing KR and ER from forming the orientation observed in the mammalian FAS.

To investigate whether the conformation of the Spn(KR+ER)2 didomain in solution is similar to that observed in the crystal structure (corrected for domain swapping), we performed SAXS (**Figures 2-2c** and **2-9**) (Putnam *et al.*, 2007). The experimental scattering curve closely matched the scattering curve predicted by the program CRYSOL for the Spn(KR+ER)2 monomer from the crystal structure (Svergun *et al.*, 1995). The scattering curve predicted from a model generated through the superposition of SpnKR2 and SpnER2 on the KR and ER domains of the porcine FAS did not match the experimentally obtained scattering curve. The radius of gyration (R_g) determined from the SAXS data through the Guinier approximation (29.5 Å) matches that calculated from the Spn(KR+ER)2 crystal structure (30.2 Å) but not that calculated from the model generated through the superposition of SpnKR2 and SpnER2 on the KR and ER domains of the porcine FAS (36.7 Å). The *ab initio* molecular envelope generated from the SAXS data by the program DAMMIF indicates a compact two-

domain structure that is consistent with the Spn(KR+ER)2 crystal structure but not with the model generated using the orientations of KR and ER in the porcine FAS (**Figures 2-2d** and **2-9**) (Svergun *et al.*, 1995; Franke & Svergun, 2009).

Oligomerization state

The architecture of PKS modules as determined by limited-proteolysis and size-exclusion experiments and through structural biology techniques has generally correlated well with the architecture of the mammalian FAS: the ketosynthase and DH are dimeric, and the acyltransferase, KR and acyl carrier protein (ACP) are monomeric (Keatinge-Clay, 2008; Keatinge-Clay & Stroud, 2006; Tang *et al.*, 2006; Alekseyev *et al.*, 2007; Maier *et al.*, 2008; Staunton *et al.*, 1996). However, studies of the oligomerization state of the ERs of modular PKSs have been seemingly contradictory: though the crystal structures of the mammalian FAS and the nucleotide-binding subdomain of the ER from PpsC show these ERs making dimeric contacts via the interface anticipated for medium-chain dehydrogenase/reductase (MDR) enzymes, limited proteolysis of DEBS2 (the second polypeptide of the erythromycin PKS) produced a monomeric EryER4 (Staunton *et al.*, 1996). A size-exclusion study of Ery(DH+KR+ER)4 suggests that this tridomain is dimeric; however, this result may not address the oligomerization state of EryER4 as it is possible that EryDH4 alone mediates dimerization (Keatinge-Clay, 2008; Keatinge-Clay & Stroud, 2006).

Spn(KR+ER)2, SpnER2 and SpnKR2 were analyzed by size-exclusion chromatography, and each migrates as a monomer (~90 kDa, ~40 kDa and ~60 kDa;

predicted values 83 kDa, 36 kDa and 50 kDa, respectively; **Figure 2-10**). The oligomerization states of these proteins were also examined through sedimentation velocity experiments (**Figure 2-11**) (Demeler & van Holde, 2004). For each Spn(KR+ER)2, SpnER2 and SpnKR2, the van Holde–Weischet integral distribution plots show essentially identical sedimentation coefficient distributions for three different concentrations examined, indicating the absence of any mass-action–based oligomerization effects. A global genetic algorithm–Monte Carlo analysis was performed to confirm the composition of and identify the molecular weights and anisotropies for each protein. The molecular weights determined from the genetic algorithm–Monte Carlo analysis are in excellent agreement with the known monomeric molecular weights of Spn(KR+ER)2, SpnER2 and SpnKR2. The molecular weight of Spn(KR+ER)2 was also estimated to be 81 kDa from the SAXS data26, providing further confirmation that Spn(KR+ER)2 is a monomer in solution.

SpnKR2 structure

SpnKR2 consists of a structural and a catalytic subdomain (residues 1–175 and 511–753, respectively; both subdomains have a Rossmann-like fold, and the KR domain belongs to the 'complex' subfamily of the short-chain dehydrogenase/reductase (SDR) superfamily), aligning well with TylKR1, AmpKR2 and EryKR1 (1.8 Å, 1.8 Å and 1.9 Å Cα r.m.s. deviation, respectively; **Figure 2-12a**) (Keatinge-Clay, 2007; Zheng & Keatinge-Clay, 2011; Keatinge-Clay & Stroud, 2006; Oppermann *et al.*, 2003). Compared to these isolated KRs, the structural subdomain of SpnKR2 is missing the first

helix and is ~30 residues shorter (Keatinge-Clay & Stroud, 2006). Within the catalytic subdomain, the active site residues Lys621, Ser643, Tyr656 and Asn660 are in the same orientations observed in other reductase-competent KRs (**Figure 2-12b**) (Keatinge-Clay, 2007; Zheng & Keatinge-Clay, 2011; Zheng *et al.*, 2010; Keatinge-Clay & Stroud, 2006). $NADP^+$ is bound in the dinucleotide-binding site of the catalytic subdomain as observed in the ternary complex of AmpKR2 (Zheng *et al.*, 2010). Adjacent to the active site, SpnKR2 has a leucine-aspartate-aspartate motif, which is used by B-type KRs to help stereoselectively generate D-hydroxyl groups, as well as a lid helix that is one turn longer than the lid helices of KRs from modules that do not contain other processing enzymes (**Figure 2-12c**) (Keatinge-Clay, 2007). In contrast to their equivalents in most other KR structures, both of these elements are well ordered and make contacts to active site residues as well as to one another, as illustrated by Arg701 of the lid helix binding the second aspartate of the leucine-aspartate-aspartate motif through a salt bridge (the interactions between these elements may have been promoted in the crystal structure by the lid helix being the first element of the C-terminal domain swap; **Figure 2-4c**).

SpnER2 structure

The ERs of modular PKSs and FASs belong to the acyl-CoA reductase family within the MDR superfamily; thus, SpnER2 has two characteristic MDR subdomains (**Figure 2-13a**) (Persson *et al.*, 2008). Residues 190–307 and 448–501 comprise the substrate-binding subdomain, whereas residues 308–447 constitute the nucleotide-binding subdomain that binds NADPH through a Rossmann fold (Maier *et al.*, 2008). The

ERs of modular PKSs are structurally distinct from their functional analogs in bacterial fatty acid biosynthesis (FabI, FabL and FabV), which are SDRs (Kim *et al.*, 2011; Lu & Tonge, 2010). SpnER2 is most structurally similar to quinone oxidoreductases (DALI server and Protein Data Bank (PDB) codes: 2J8Z, 3JYN and 1QOR; 2.1 Å, 2.3 Å and 2.4 Å Cα r.m.s. deviation, respectively) (Holm & Rosenstrom, 2010). SpnER2 is also highly structurally related to the ER of the mammalian FAS (PDB code 2VZ8; 2.1 Å Cα r.m.s. deviation), a *trans*-acting enoylreductase in the lovastatin biosynthetic pathway (PDB code 3B70; 2.6 Å Cα r.m.s. deviation), and 2-enoyl thioester reductases from mitochondrial fatty acid synthesis pathways (PDB codes 2VCY, 1N9G and 1GUF; 3.1 Å, 3.7 Å and 3.7 Å Cα r.m.s. deviation, respectively) (Maier *et al.*, 2008; Chen *et al.*, 2008; Torkko *et al.*, 2003; Airenne *et al.*, 2003). The boundaries of the SpnER2 domain are defined by two β-strands (β1 and β9) in the substrate-binding subdomain. The substrate-binding and nucleotide-binding subdomains are oriented relative to one another largely by helix αA (secondary-structure elements of the substrate-binding subdomain are labeled with numbers, and those of the nucleotide-binding subdomain are labeled with letters, as is customary for the Rossmann fold), which spans both subdomains and is bent through the insertion of a side chain carboxylate (from Glu250) at its third turn. The nucleotide-binding subdomain of SpnER2 adopts a classical Rossmann fold, whereas the nine β-strands of the substrate-binding subdomain form two β-sheets (β1 and β2 create a small, antiparallel β-sheet, and β3–β9 build a highly twisted, mixed β-sheet) surrounded by four α-helices (α1, αA, α2 and α3).

MDRs usually dimerize via βF and the secondary structural elements around it (Persson *et al.*, 2008). The surrounding elements in the ER from the mammalian FAS are αF, βG and αG (Kirst, 2010). SpnER2 has αF but not βG or αG. Quinone oxidoreductases share a similar one-helix architecture to SpnER2; however, interactions between Asp405 and Arg426 as well as Leu408 and His435 position helix αF of SpnER2 in an orientation that is seemingly incompatible with dimerization (**Figures 2-13b**, **2-7**, **2-8**, and **2-14a**) (Thorn *et al.*, 1995). The highly related *trans*-acting lovastatin enoylreductase (PDB code 3B70) is the only other MDR enzyme that crystallized as a monomer.

The NADPH-binding site is located in a deep cleft between the subdomains of SpnER2 (**Figures 2-13c** and **2-14b**,**c**). The pyrophosphate moiety of the nicotinamide coenzyme is bound to the highly conserved pyrophosphate-binding motif GGVGMA (residues 335–340; **Figure 2-7**) (Maier *et al.*, 2008). The conserved residues Lys360 and Arg375 form salt bridges with the adenine ribose phosphate. Elements from the substrate-binding subdomain, such as Phe231 from helix α1 and His494 from the loop following helix α3, also interact with the bound nicotinamide coenzyme. Thus, the 4-pro-*R* hydride of NADPH is positioned to add to the β-carbon of a *trans*-α,β-unsaturated polyketide substrate, in contrast to KRs, which catalyze the addition of the 4-pro-*S* hydride (Zheng *et al.*, 2010).

SpnER2 function

The structures of both the ER from the porcine FAS and SpnER2 reveal an invariant lysine-aspartate pair (Lys422 and Asp444 in SpnER2) ~6 Å from the

nicotinamide hydride (Maier *et al.*, 2008). They may directly participate in substrate protonation, stabilize reaction intermediates or aid in binding the substrate (**Figure 2-13c**). Structural and functional analyses of many MDRs suggest that a tyrosine following helix α1 is the proton source during the reduction reaction; however, only half of the ERs from modular PKSs have the corresponding tyrosine (**Figure 2-7**). When present, it is known to have the major role in setting an *S* configuration at the α-carbons of α-substituted polyketide intermediates (Kwan *et al.*, 2008).

The activities and stereocontrol of isolated KRs have been extensively analyzed *in vitro*34–37, whereas all functional assays of ER domains have been limited to *in vivo* assays within the context of complete modules (Kwan *et al.*, 2008; Kwan & Leadlay, 2010). Thus, Spn(KR+ER)2 was assayed for *in vitro* activity toward the *trans*-α,β-unsaturated crotonyl-pantetheine, which mimics its natural substrate, *trans*-(5*S*)-hydroxyhept-2-enoyl-SpnACP2 thioester (**Figures 2-1b** and **2-15**). Indeed, Spn(KR+ER)2 generates the expected product, butyryl-pantetheine (verified by MS; **Figure 2-16**). A recent study attempted to determine the catalytic residues of the ERs of modular PKSs using an *in vivo* assay; however, activity was preserved in each of the generated point mutants (Kwan & Leadlay, 2010). We sought to identify the catalytic residues of ER through an *in vitro* assay with crotonyl-pantetheine. The structures of the mammalian FAS and Spn(KR+ER)2 and previous studies guided the selection of residues to be examined (Kwan *et al.*, 2008; Kwan & Leadlay, 2010; Maier *et al.*, 2008). Y241F, K422A and S444A point mutants were expressed and purified using the same protocol as described for wild-type Spn(KR+ER)2. Along with the anticipated butyryl-pantetheine

product, a side product with the molecular weight of β-hydroxybutyryl–pantetheine was generated (**Figure 2-15**). That the K422A mutant produced the least butyryl-pantetheine suggests that the lysine is most critical for the reductase activity of SpnER2.

We sought to determine the steady-state kinetic parameters of the reduction reaction catalyzed by SpnER2 and its point mutants on the substrate crotonyl-pantetheine (**Table 2-2**). Because we observed that stand-alone SpnKR2 could catalyze the hydration of crotonyl-pantetheine, the Y241F, K422A and D444A mutations were introduced into stand-alone SpnER2. The catalytic efficiency of the reduction reaction (R) of wild-type SpnER2 (k_{cat}(R)/K_m = 0.11 ± 0.02 M−1 s−1; k_{cat}(R) = 0.0085 ± 0.0007 s−1 and K_m = 74 ± 14 mM) is similar to those of isolated KR domains for diketide-*N*-acetylcysteamine thioesters (for example, EryKR1: k_{cat}/K_m = 7 M−1 s−1; k_{cat} = 0.26 ± 0.01 s−1 and K_m = 35 ± 4 mM)35 but much lower than that of the ER domain of the intact *Rattus norvegicus* FAS for crotonyl-CoA (k_{cat}/K_m = 1.6 × 106 M−1s−1; k_{cat} = 9.8 ± 0.5 s−1 and K_m = 6.5 ± 0.6 μM)38. The D444A mutant was slightly more catalytically efficient than wild-type SpnER2 (k_{cat}(R)/K_m = 0.30 ± 0.04 M^{-1} s^{-1}). The Y241F mutant was less efficient (k_{cat}(R)/K_m = 0.08 ± 0.01 M−1 s−1) than the wild type, whereas the K422A mutant was the least catalytically efficient of the mutants (k_{cat}(R)/K_m = 0.018 ± 0.004 M−1 s−1). Each of the SpnER2 point mutants more actively catalyzed the hydration side reaction than wild-type SpnER2 (**Table 2-3**).

DISCUSSION

The order of PKS enzymes within a module and the reactions they catalyze closely parallel those of the mammalian FAS (Khosla *et al.*, 2007; Smith & Tsai, 2007). Nonetheless, the combinations of processing enzymes present within PKS modules enable a much wider range of products to be generated. Determining the architectures and mechanisms of these enzymes are necessary steps toward elucidating how these complicated synthetic machines generate chemically diverse polyketides. This structural and functional report of a KR-ER didomain from a complete PKS module represents a large step toward this goal.

Although no mechanism for the ERs from mammalian FASs or modular PKSs has been generally accepted, it is thought that NADPH transfers a hydride to the β-carbon of the substrate and that a proton is donated by either a general acid or a solvent molecule to the α-carbon (**Figures 2-13c** and **2-15**) (Smith & Tsai, 2007; Maier *et al.*, 2008). That Lys422 was found to be the most crucial for catalysis makes sense as it is structurally conserved in the ERs of both mammalian FASs and modular PKSs, whereas the tyrosine is absent in the ERs from mammalian FASs (and half of the ERs from modular PKSs) and the aspartate resides in a different location within the ER of the mammalian FAS (**Figures 2-7** and **2-8**) (Maier *et al.*, 2008).

The mechanisms by which some ERs from modular PKSs set stereochemistry are made less mysterious through the structure of SpnER2. The tyrosine that has the principal role in setting an *S* configuration at a substituted α-carbon (equivalent to Tyr241 in SpnER2) lies on the opposite side of the active site cleft from the lysine (Maier *et al.*,

2008). The lysine and tyrosine are within 5 Å of one another, and both residues are in appropriate positions to protonate the α-carbon of a bound polyketide substrate during the reduction reaction (**Figure 2-13c**). Mutation of this tyrosine has been observed to result in an *R* configuration being set at the α-carbon of an α-substituted substrate, suggesting that, in the absence of the tyrosine, the lysine is able to donate a proton from the opposite side of the polyketide substrate (Kwan *et al.*, 2008). As shown in the mutagenesis results presented here, the lysine serves a key role even when the tyrosine is present, possibly lowering the pK_a of the tyrosine hydroxyl group so that it may readily donate a proton to the substrate. We have also observed the reduction of 2-methylbut-2-enoyl-pantetheine by SpnER2; how stereocontrol is enforced by the ERs of modular PKSs on this substrate is under investigation.

The structures of the monomeric thioesterase from the mammalian FAS and the dimeric thioesterase from multimodular PKSs revealed that structural differences exist between these related megasynthases (Tsai *et al.*, 2008; Chakravarty *et al.*, 2004). The monomeric state of SpnER2 also speaks to the divergent evolution of multimodular PKSs and mammalian FASs. Their common ancestor most likely had a dimeric ER; however, when some PKSs evolved so that a single polypeptide contained multiple modules, disruption of the ER dimer interface may have been necessary to enable the ACPs, restricted by a dimeric element immediately downstream, to access each of their cognate enzymes. The DH dimer interface within a multimodular PKS is quite different from that of the mammalian FAS; as with an evolutionary transition to a monomeric ER, a more flattened DH dimer in multimodular PKSs may have facilitated the access of ACPs to

their cognate enzymes (Keatinge-Clay, 2008; Akey *et al.*, 2010). No structural information is available for how a DH domain in a multimodular PKS structurally interacts with KR and ER; the presence of DH could result in KR and ER adopting an interface different from that observed in the Spn(KR+ER)2 crystal structure. The homodimeric oligomerization state of the nucleotide-binding subdomain of the ER from phthiocerol synthase PpsC seems to argue against an evolutionary transition to monomeric ERs within multimodular PKSs (PDB code 1PQW; the interface is equivalent to that of the ER in the mammalian FAS) (Pedelacq *et al.*, 2011). However, PpsC is more related to mammalian FASs on a sequence level than modules from multimodular PKSs—the C-terminal linker as well as the sequence surrounding βF are very similar to that of the mammalian FAS (**Figures 2-13b**, **2-6**, **2-7**, and **2-8**) (Maier *et al.*, 2008). Perhaps the architecture of some monomodular PKSs such as PpsC resembles that of the mammalian FASs more than that of the modules of multimodular PKSs.

Models of complete PKS modules have been constructed on the basis of the structure of the mammalian FAS (Khosla *et al.*, 2007; Maier *et al.*, 2008). However, the KR-ER interface observed in Spn(KR+ER)2, the short C-terminal linker between ER and KR, the altered tertiary structure of the MDR dimerization region and the flattened architecture of the DH dimer of multimodular PKSs indicate that the three-dimensional organization of the processing enzymes of multimodular PKSs is quite different from that of the mammalian FAS (**Figures 2-2**, **2-6**, **2-7**, **2-8**, **2-9**, and **2-10**) (Keatinge-Clay, 2008; Maier *et al.*, 2008). Indeed, the two-fold axis of a multimodular PKS does not contain enough space for ER domains to form dimeric contact as the ACPs, geometrically

restricted by ~20-residue linkers to KR and a downstream dimeric element, require this space to access enzymes from the same module (the mammalian FAS and some monomodular PKSs do not encounter this geometric problem because their ACPs are not C-terminally linked to a dimeric domain). Thus, we propose that within multimodular PKSs the 'reductive loop' comprises a dimeric DH, two monomeric KRs and two monomeric ERs, thereby granting the ACPs more freedom to access cognate enzymes and preserving the two-fold symmetry of the synthase (**Figure 2-17**). Although the crystal structure of a complete PKS module remains elusive, the structural and functional dissection of its component enzymes provides insights into the architectures and mechanisms of these complex megasynthases.

MATERIALS AND METHODS

Cloning, expression, and purification

Domain boundaries from previous KR structures guided the selection of the first and last residues of the Spn(KR+ER)2 didomain (SpnB residues 1216-1989) (Waldron *et al.*, 2001). The DNA encoding Spn(KR+ER)2 was amplified using primers 5'-ATCGTAATC*CATATG*GCTCGGTTGTACCGGTTGAG-3' and 5'-TGATTCGAT*GAATTC*ACAACCGTTTCCGGAGGCTCT-3' (NdeI and EcoRI sites in italics; stop codon underlined) from *Saccharopolyspora spinosa* genomic DNA (extracted from *S. spinosa* using the DNeasy Blood & Tissue Kit from Qiagen), digested with NdeI and EcoRI, and ligated into pET28b (Novagen). The DNA encoding the SpnER2 domain was amplified using primers 5'-

ATCGTAATC*CATATG*GGCACCGGGTGGCGATTGGA-3' and 5'-TGATTCGAT*GAATTC*ACACCGGAGGCATGGTGAGCA-3' and ligated into pET28b (Novagen) as described above. Reverse PCR was used to remove the ER domain from the Spn(KR+ER)2 didomain construct using 5'-phosphorylated primers 5'-CCTCCGGTGTGGGACGCCGCAGGCACGGTTCTGGTTAC-3' and 5'-CCCGGTGCCGTCCGGCACGGGCAGGACGTCATCGCT-3'. *E. coli* BL21(DE3) transformed with the expression plasmid was inoculated into LB media containing 50 mg/L kanamycin at 37 °C, grown to OD600 = 0.4, and induced with 0.5 mM IPTG. After 12 h at 15 °C, cells were collected by centrifugation and resuspended in lysis buffer (300 mM NaCl, 50 mM HEPES, pH 7.5). Following sonication, cell debris was removed by centrifugation (30,000 x g for 45 min). The supernatant was poured over a column of Nickel-NTA resin (Qiagen), which was then washed with 50 mL lysis buffer containing 15 mM imidazole and eluted with 5 mL lysis buffer containing 300 mM imidazole. Proteins were further purified using a Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated with 150 mM NaCl, 10 mM HEPES, pH 7.5. Spn(KR+ER)2 was concentrated to 24 mg/mL in 25 mM NaCl, 1 mM DTT, 10 mM HEPES, pH 7.5. Selenomethionine-labeled Spn(KR+ER)2 was obtained by the pathway inhibition method (Keatinge-Clay & Stroud, 2006). Briefly, BL21(DE3) cells were grown in M9 minimal media containing 4 g/L glucose, 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 50 mg/L kanamycin at 37 °C. Lysine, phenylalanine, threonine (each at 100 mg/L), isoleucine, leucine, valine (each at 50 mg/L), and selenomethionine (at 60 mg/L) were supplied at OD600 = 0.4. After 15

minutes at 37 °C, 0.5 mM IPTG was added, and the cells were induced for 16 hours at 15 ºC. The purification followed the protocol described for the native protein.

Site-directed mutagenesis

The GeneTailor Site-Directed Mutagenesis System (Invitrogen) was used to generate all the mutants (verified by sequencing). Purification of the point mutants followed the protocol described for the wild-type protein. The following oligonucleotides were used for mutagenesis: K422A, 5'-GCCGTTTCCTGGAGTTGGGGGCGACGGATGTTCGTGACC-3' and 5'-CCCCAACTCCAGGAAACGGCCACCGCGCGGCAGCATTC-3'; D444A, 5'-CGTGTCTTACCAGGCTTTCGCTACCGTAGAGGCAGGCC-3', and 5'-CGAAAGCCTGGTAAGACACGCCCGGATGCGCATCGGC-3'; Y241F, 5'-GCTCATCGCGCTCGGTATGTTTCCCGGTGTGGCATCGCT-3' and 5'-ACATACCGAGCGCGATGAGCGCATCCCGGAAGTTCAC-3'.

Size-exclusion chromatography

Each sample (0.3 mL; 69 μM Spn(KR+ER)2; 132 μM SpnER2; 75 μM SpnKR2) was injected onto a Superdex 200 gel-filtration column equilibrated with 150 mM NaCl, 10 mM sodium phosphate, pH 7.4. The molecular weight was estimated by comparison with standards (Gel Filtration Standard, Biorad Laboratories) as previously described (Adams *et al.*, 2010).

Analytical ultracentrifugation

All sedimentation experiments were performed with a Beckman Optima XL-I at the Center for Analytical Ultracentrifugation of Macromolecular Assemblies at the University of Texas Health Science Center at San Antonio (UTHSCSA). Sedimentation velocity data were analyzed with the UltraScan-II software version 9.93. All calculations were performed on the Lonestar and Ranger clusters at the Texas Advanced Computing Center at the University of Texas at Austin, and on the Jacinto cluster at the Bioinformatics Core Facility at UTHSCSA. All measurements were made at 280 nm in a buffer containing 150 mM NaCl, 10 mM sodium phosphate, pH 7.4. The experimental data were collected at 20 °C and at 50 krpm using standard Epon 2-channel centerpieces. Hydrodynamic corrections for buffer density, viscosity, and partial specific volume were made according to methods implemented in UltraScan (Otwinowski *et al.*, 1997). The partial specific volumes of Spn(KR+ER)2, SpnER2, and SpnKR2 were determined to be 0.737 cm3/g, 0.736 cm3/g, and 0.735 cm3/g, respectively. All data were first analyzed by 2-dimensional spectrum analysis (2DSA) with simultaneous removal of time- and radial-invariant noise and fitting of the meniscus position, followed by van Holde-Weischet analysis and genetic algorithm (GA) refinement, followed by Monte Carlo analysis (MC) (Brookes *et al.*, 2009; Demeler & Brooks, 2008). Each protein was measured at 280 nm at three loading concentrations (Spn(KR+ER)2, 1.2 μM, 4.1 μM, and 12.0 μM; SpnER2, 3.0 μM, 10.5 μM, and 31 μM; SpnKR2, 2.5 μM, 7.5 μM, and 23 μM).

Crystallization and structure determination

Crystals of Spn(KR+ER)2 grew over one week by sitting drop vapor diffusion at 22 °C. Drops were formed by mixing 2 μL protein solution (24 mg/mL Spn(KR+ER)2 in 25 mM NaCl, 1 mM DTT, 10 mM HEPES, pH 7.5) with 1 μL crystallization buffer (17% (w/v) poly(acrylic acid) 5100, 0.4 M MgSO4, 0.1 M HEPES, pH 6.5). Crystals were frozen in liquid nitrogen. Three data sets (peak, inflection, and remote) from a crystal of selenomethionine-labeled Spn(KR+ER)2 were collected at ALS Beamline 5.0.2 and processed by HKL2000 (Otwinowski & Minor, 2007). The structure was solved to 3.40 Å resolution by multiwavelength anomalous dispersion phasing using the program Phenix (Adams *et al.*, 2010). The program Molrep fit the resulting electron density map with AmpKR2 (PDB Code: 3MJS) and the ER from the porcine FAS (PDB Code: 2ZV8), and the solutions were combined to obtain the model of two Spn(KR+ER)2 molecules in the asymmetric unit (Potterton *et al.*, 2003; Maier *et al.*, 2008; Zheng *et al.*, 2010). A 3.0 Å-resolution data set collected from a crystal of native Spn(KR+ER)2 at ALS Beamline 8.2.1 was used to iteratively build and refine the model through the programs Coot and Refmac5 (Potterton *et al.*, 2003; Emsley & Cowtan, 2004).

SAXS data collection and analysis

Spn(KR+ER)2 was exchanged by gel filtration chromatography (Superdex 200) into 10% glycerol, 500 mM NaCl, 50mM $(NH_4)_2SO_4$, and 30 mM HEPES, pH 7.5. SAXS data were collected at 5 °C on a Rigaku BioSAXS-1000 (Kratky camera, with a 2.5kW FRE+ source) at the Sealy Center for Structural and Molecular Biophysics (SCSB)

BioSAXS facility of the University of Texas Medical Branch at Galveston. A dilution series showed no difference between a 10 mg/mL and a 5 mg/mL sample; therefore, a sample concentration of 10 mg/mL was used for data collection. Five 30-minute exposures were collected. Radiation-induced aggregation effects slowly developed in the low-q regime (q<1.3/R_g); however, the scattering curves of all samples converged beyond this point, permitting all of the curves to be included in the averaging for the high-q regime (q>1.3/R_g). Five more 30-minute data collections were performed, each with fresh sample. Using the SAXSLab software (Rigaku), the high-q regions of these ten collections were averaged, and the low q-regime of the last five collections were also averaged. The high- and low-q curves were merged together and scattering from the buffer alone was subtracted (seven 30-minute scans were collected on the buffer alone) with the program PRIMUS (Konarev *et al.*, 2003). The resulting SAXS curve was used for subsequent analysis. The radius of gyration (R_g) and $I(0)$ were evaluated using the Guinier approximation (0.0223 $Å^{-1}$ < q < 0.0437 $Å^{-1}$). The program GNOM was used to obtain the distance distribution function, $P(r)$, and to determine the value for the R_g from the entire scattering profile (Svergun, 1992). Good agreement of R_g(Guinier) and R_g(GNOM) indicated the internal consistency of the data (Putnam *et al.*, 2007). A molecular weight of 81 kDa was calculated from the SAXS curve using the invariant Kratky integral method (Fischer *et al.*, 2010). The program CRYSOL predicted the scattering curves from two models (one model of Spn(KR+ER)2 based on its crystal structure and one model of Spn(KR+ER)2 based on the superposition of SpnKR2 and SpnER2 on the KR and ER domains of the porcine FAS) and compared them to the

experimentally-obtained SAXS curve (Svergun *et al.*, 1995). *Ab initio* envelopes were generated by the program DAMMIF (Franke & Svergun, 2009). A developmental version of the pmb_dammif script within the Protein Model Builder (PMB) package was used to generate fifty envelopes, which were averaged and filtered with the program DAMAVER (Volkov & Svergun, 2003). The resulting consensus envelope was then aligned and fit to each model using the program SUPCOMB (Kozin & Svergun, 2000).

Substrate synthesis

Pantethine (278 mg, 0.5 mmol, 1 eq.) was incubated with DTT (77 mg, 0.5 mmol, 1 eq.) in 1 mL water and stirred for 1 hour at room temperature. The water was removed under vacuum, and the clear oil was resuspended in THF (8 mL) and triethylamine (2 mL). Dimethylaminopyridine (6 mg, 0.05 mmol, 0.1 eq.) was added, and the solution was stirred on ice. Chilled crotonyl chloride (50 µL, 55 mg, 0.525 mmol, 1.05 eq.) was added, and the reaction was stirred for one hour. The reaction was allowed to warm to room temperature and was stirred for another hour. Solvent was removed under vacuum, and the remaining orange residue was resuspended in 50% ethyl acetate/acetone. Using the same solvent system as a mobile phase, crotonyl-pantetheine was purified by silica gel chromatography. Fractions containing crotonyl-pantetheine (R_f = 0.30, silica TLC plate, EtOAc) were dried under vacuum; the clear residue was resuspended in water and the solution was frozen until needed.

Crotonyl-pantetheine. Electrospray ionization mass spectrometry (ESI-MS), expected mass 347.4; observed mass: [M+H]+ 347.2 (also observed [M+H-H_2O]+, 329.2; and

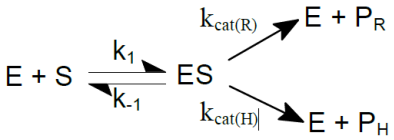
[M+Na]+, 369.4). A Varian 400-MHz DirectDrive instrument was used to obtain NMR spectra. ^{1}H NMR (D2O, 400 MHz) δ 7.05-6.95 (dq, 1H, J=15.53, 6.91); 6.29-6.24 (dq, 1H, J=15.53, 1.68); 3.95 (s, 1H); 3.51-3.32 (m); 3.07 (t, 2H, J=6.30); 2.44 (t, 2H, J=6.49); 1.87 (dd, 3H, J=6.88, 1.69); 0.89 (s, 3H); 0.85 (s, 3H). These peaks match those previously reported (Gomes *et al.*, 1991).

HPLC and LC/MS characterization of Spn(KR+ER)2 functional assays

To assay the activity of isolated Spn(KR+ER)2 and its point mutants, the following reaction was initiated: 40 μM Spn(KR+ER)2, 2 mM crotonyl-pantetheine, 1 mM $NADP^{+}$, 150 mM glucose, 3 μM glucose dehydrogenase, 12% (v/v) glycerol, 150 mM HEPES, pH 7.0. After 48 hours, the reaction was injected on a C18 reversed-phase HPLC column, and monitored at 235 nm (100% water to 100% MeOH, 30 min.). For LC/MS analysis, an Agilent Technologies system (1200 Series HPLC, 6130 quadrupole mass spectrometer) was employed; a gradient of 5-95% B (solvent A, water with 0.1% formic acid; solvent B, acetonitrile with 0.1% formic acid) was run using a Gemini C18 column (5 μm, 2 x 50 mm) over 12 min. at a flow rate of 0.7 mL/min. Both the crotonyl-pantetheine substrate and the butyryl-pantetheine product have been previously synthesized and characterized (Gomes *et al.*, 1991; Tran *et al.*, 2008).

Kinetic assays

A kinetic model derived from standard Michaelis–Menten kinetics was employed for the reduction and hydration reactions catalyzed by SpnER2.



(1)

$$K_m = \frac{k_{-1}+k_{cat(R)}+k_{cat(H)}}{k_1} \qquad (2)$$

$$V_{0(R)} = \frac{k_{cat(R)}[E_T][S]}{K_m+[S]} \qquad (3)$$

Each reaction was performed in duplicate (50 µL reaction volume: 2.5-160 mM crotonyl-pantetheine, 1 mM NADPH, 20 µM enzyme, 10% glycerol, 100 mM NaCl, 100 mM HEPES, pH 7.5 at 22 ºC). The consumption of NADPH was monitored at 340 nm, and readings were corrected for the nonenzymatic consumption of NADPH. Initial velocities were plotted against substrate concentration and fit to the kinetic model to obtain K_m and k_{cat}(R) using the program Prism. To compare the hydration/reduction activities of SpnER and its point mutants the ratio of hydrated to reduced product generated by each SpnER2 mutant was compared to that of SpnER2. Reactions were stopped through the addition of an equal volume of methanol and analyzed by LC/MS (after removing precipitated protein by centrifugation). Extracted ion chromatograms for the hydration product (365.2 Da) and the reduction product (349.2 Da) were used to obtain the peak areas for the hydrated and reduced products, respectively.

ACKNOWLEDGEMENTS

We thank A.F. Monzingo for helping with in-house diffraction experiments and C.D. Fage for helping optimize crystals. Synchrotron data were obtained at the ALS

Beamlines 8.2.1 and 5.0.2. Financial support was provided by Welch Foundation Grant F-1712 (A.T.K.-C.) as well as the Sealy and Smith Foundation grant to the Sealy Center for Structural Biology and Molecular Biophysics (M.A.W.). The development of the UltraScan software is supported by the US National Institutes of Health through grant RR022200 (B.D.). Supercomputer time allocations were provided through the US National Science Foundation grant TG-MCB070038 (B.D.). We acknowledge the support of the San Antonio Cancer Institute grant P30 CA054174 for the Center for Analytical Ultracentrifugation of Macromolecular Assemblies at the University of Texas Health Science Center at San Antonio.

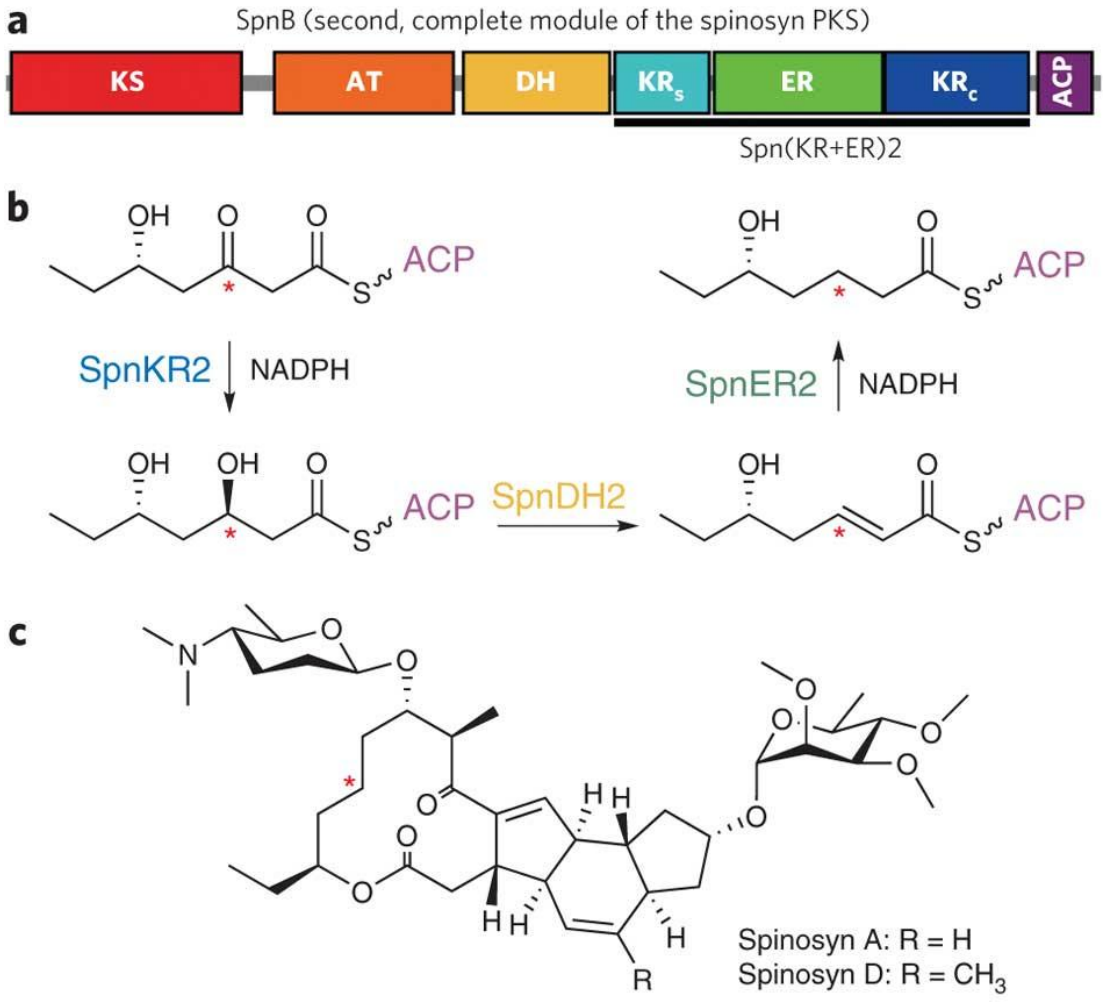


Figure 2-1. A complete module from the spinosyn PKS. **(a)** Domain composition of the second module of the spinosyn PKS (SpnMod2) and location of the Spn(KR+ER)2 didomain (approximate sequence scale). KS, ketosynthase; AT, acyltransferase; KR_s, ketoreductase structural subdomain; ER, enoylreductase; KR_c, ketoreductase catalytic subdomain; ACP, acyl carrier protein. **(b)** The processing enzymes of SpnMod2 reduce the carbon indicated by an asterisk in the polyketide intermediate to a methylene group that appears in the spinosyn products. **(c)** The complex polyketides spinosyn A and spinosyn D comprise the insecticide spinosad.

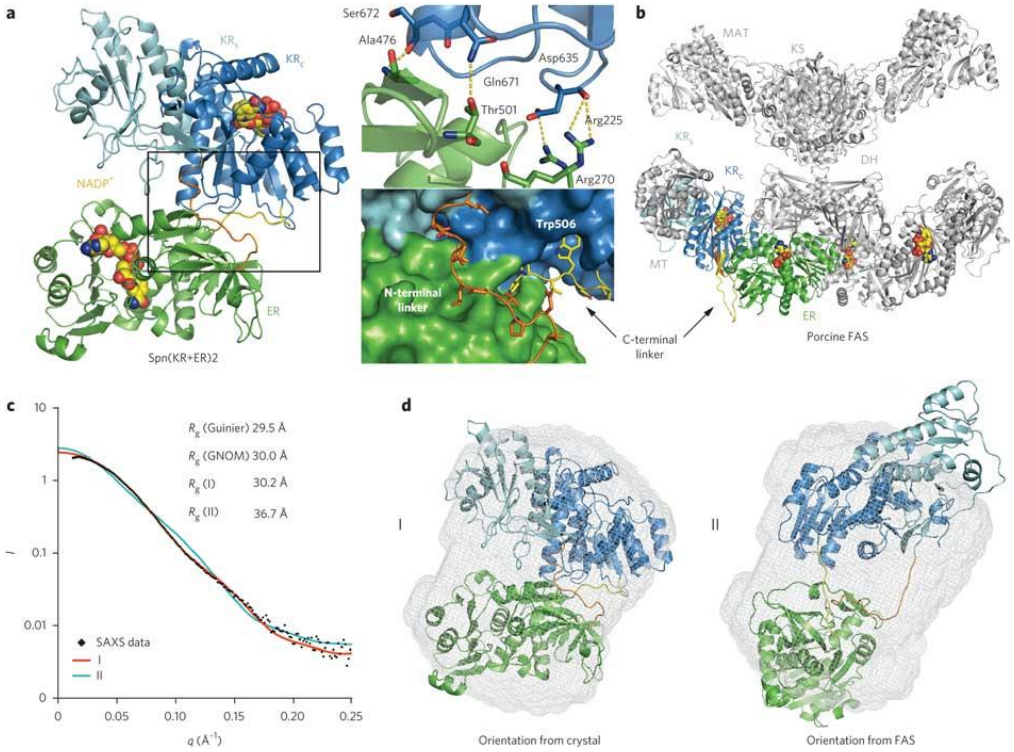


Figure 2-2. KR and ER architecture. **(a)** The crystal structure of Spn(KR+ER)2 shows the interface between the structural and catalytic subdomains of its KR (KR_s and KR_c, respectively) and ER, which is mostly formed by hydrophilic residues. The N- and C-terminal linkers help mediate the interface, Trp506 inserting into a hydrophobic pocket of KR_c. **(b)** The structure of the porcine FAS shows a very different interface between its KR and ER and how ER dimerizes across the two-fold axis of the FAS. The C-terminal linker is longer than that of Spn(KR+ER)2 (29 versus 8 residues). KS, ketosynthase; MAT, malonyl/acetyl-CoA transacylase; MT, methyltransferase. **(c)** SAXS data were obtained for Spn(KR+ER)2 and compared to the theoretical curves predicted by the program CRYSOL for a model generated from the Spn(KR+ER)2 crystal structure (red curve) and for a model generated by superposing SpnKR2 and SpnER2 on the KR and ER from the porcine FAS (blue curve). I and II refer to the orientations shown in panel d. **(d)** An *ab initio* molecular envelope was calculated from the SAXS curve (by the program DAMMIF), which was much better fit by the model generated from the Spn(KR+ER)2 crystal structure than the model generated by superposing SpnKR2 and SpnER2 on the KR and ER of the porcine FAS (using the program SUPCOMB).

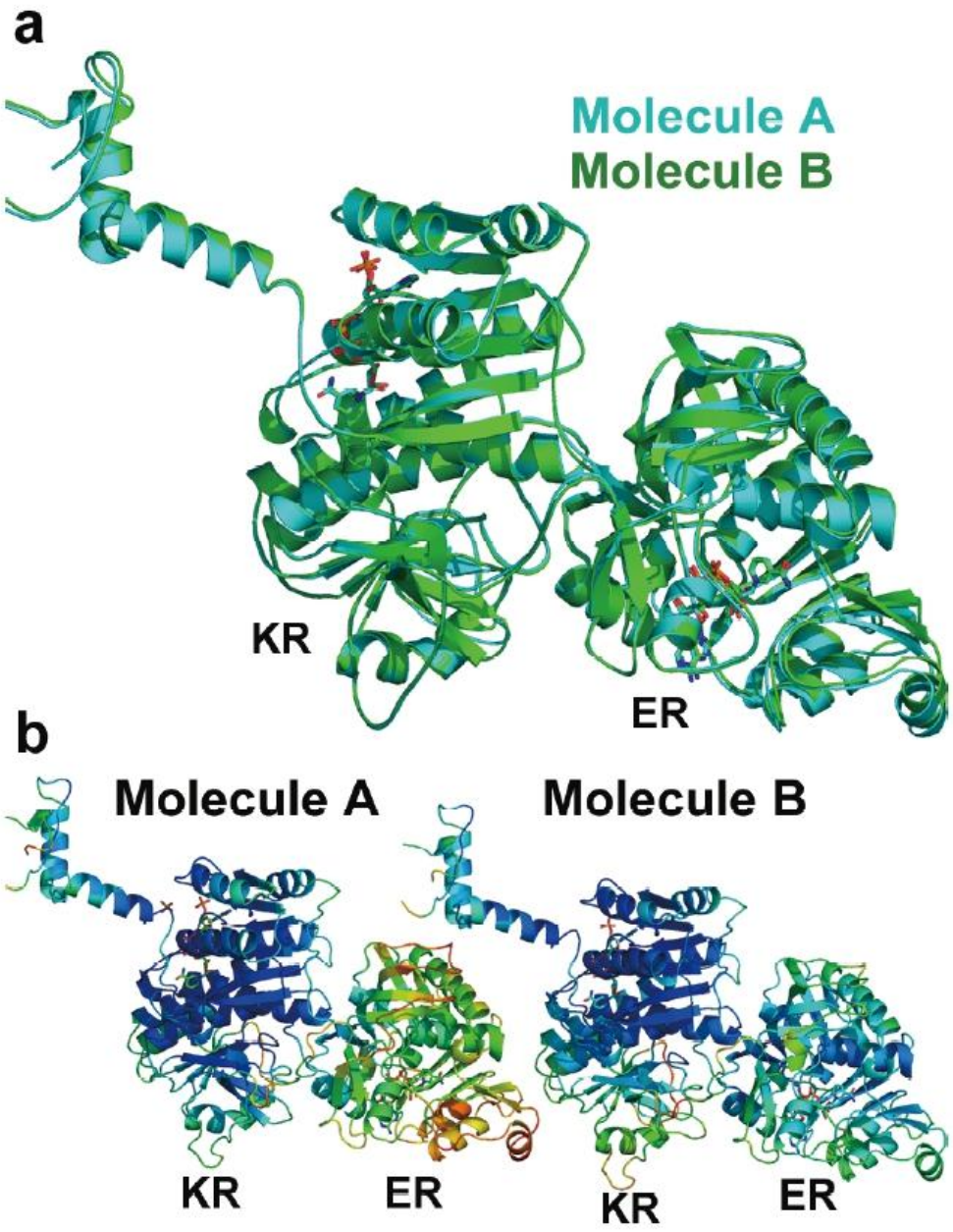


Figure 2-3. Comparison of the two Spn(KR+ER)2 molecules in the asymmetric unit. **(a)** The superposition of the Spn(KR+ER)2 didomains shows their near-identical conformations (0.6 Å Cα r.m.s.d.). **(b)** The B-factors of ER in Molecule A are elevated compared to those of ER in Molecule B (96 $Å^2$ vs. 62 $Å^2$) (B-factors are indicated on a scale from blue to red).

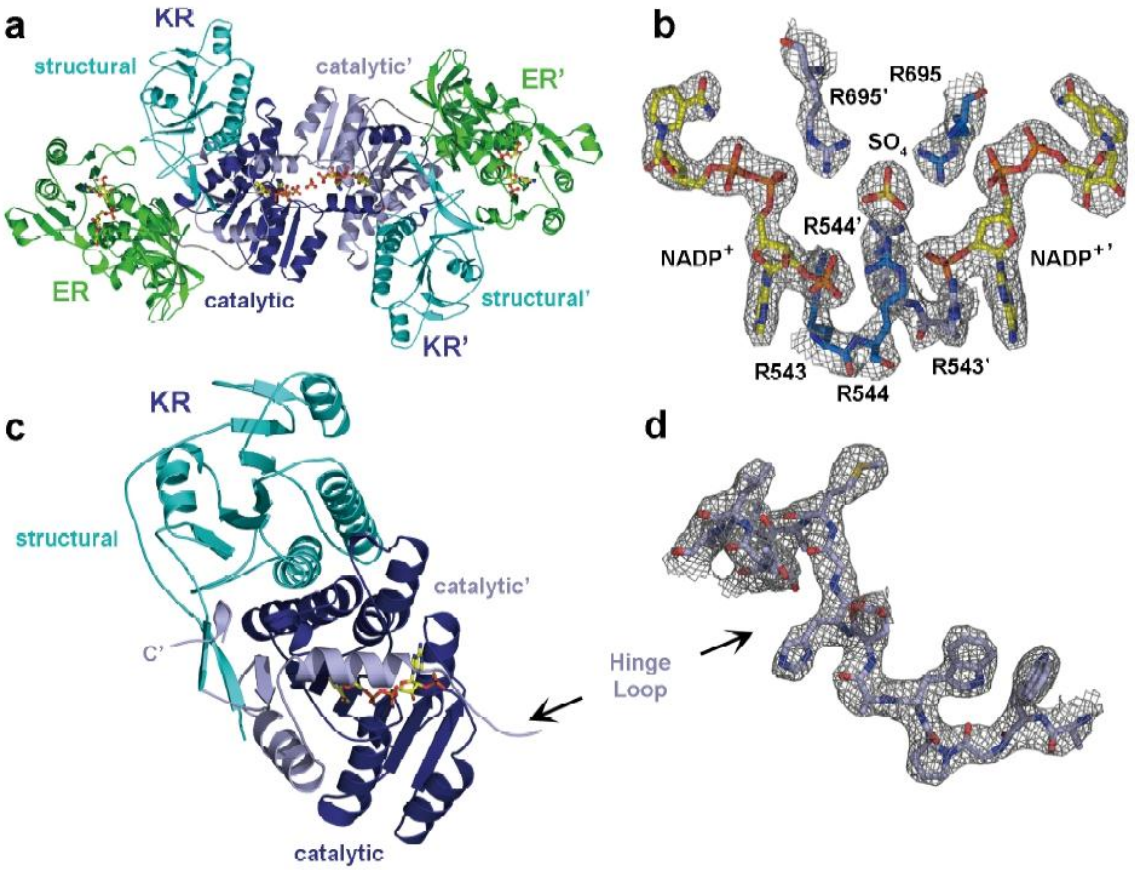


Figure 2-4. Domain-swapped, C-terminal segment. **(a)** Within the asymmetric unit, a KR/KR interface is formed between the two Spn(KR+ER)2 molecules. **(b)** The $2F_o$-F_c electron density map (contoured at 1.7 σ) shows the ionic interactions between arginines, $NADP^+$ molecules, and a sulfate at the pseudo-twofold axis of the asymmetric unit. **(c)** The 60 residues from the adjacent Spn(KR+ER)2 molecule are colored in grey. The positions of the swapped residues are as expected from related KR structures. **(d)** The $2F_o$-F_c electron density map (contoured at 1.7 σ) shows the well-ordered hinge loop (terminology from domain-swapping literature) that spans the Spn(KR+ER)2 molecules. The lid helix on the left and the tryptophans on the right are located within separate Spn(KR+ER)2 molecules, but where they would be expected from related KR structures.

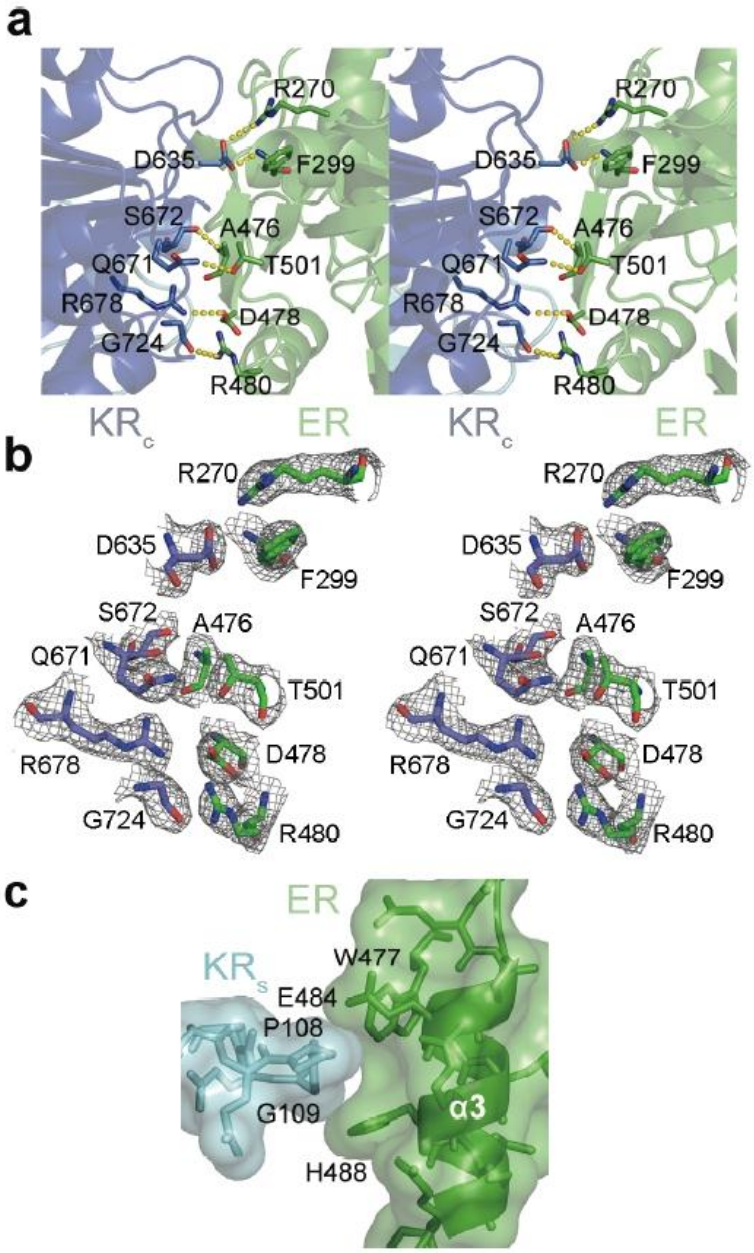
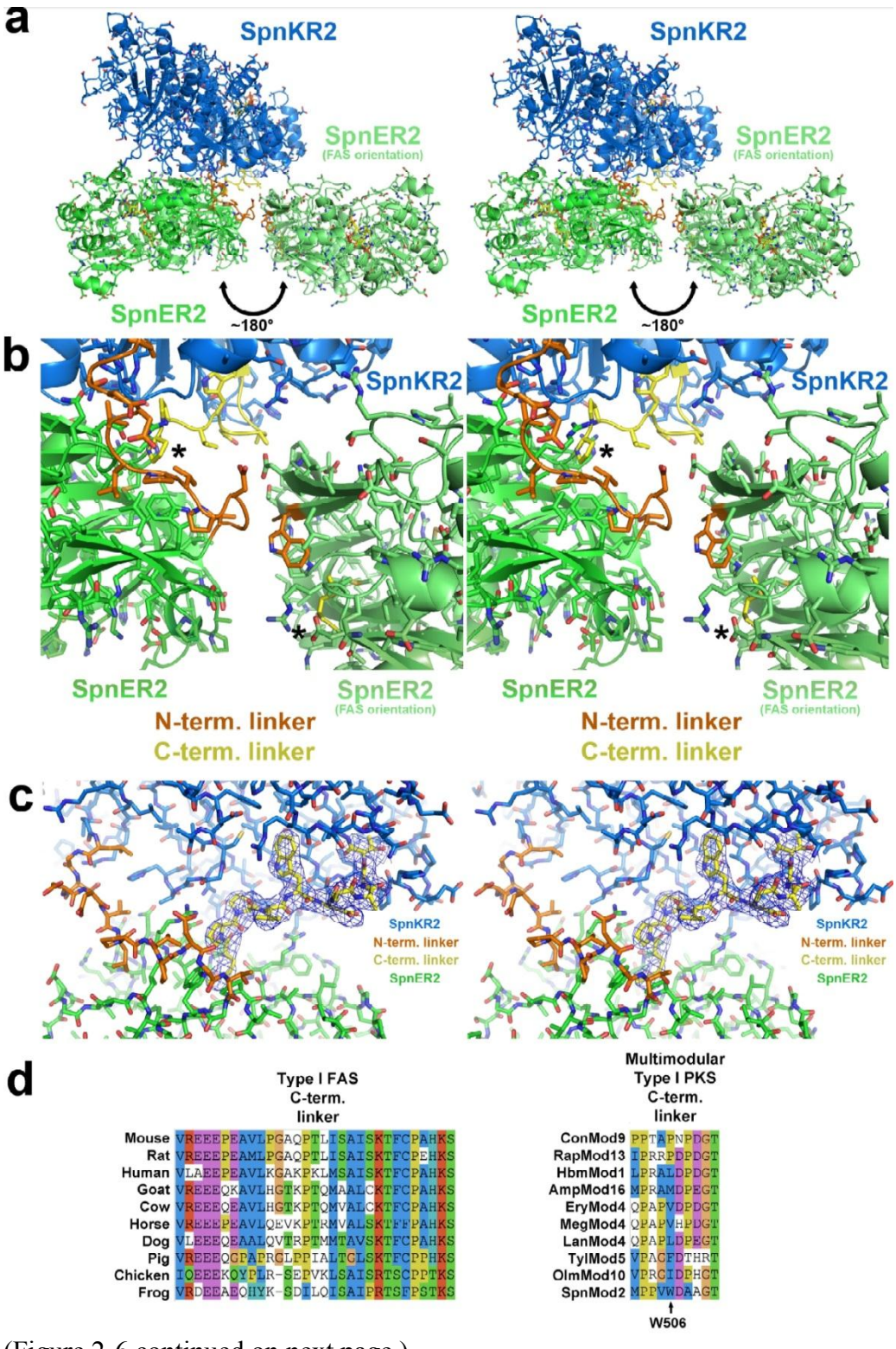
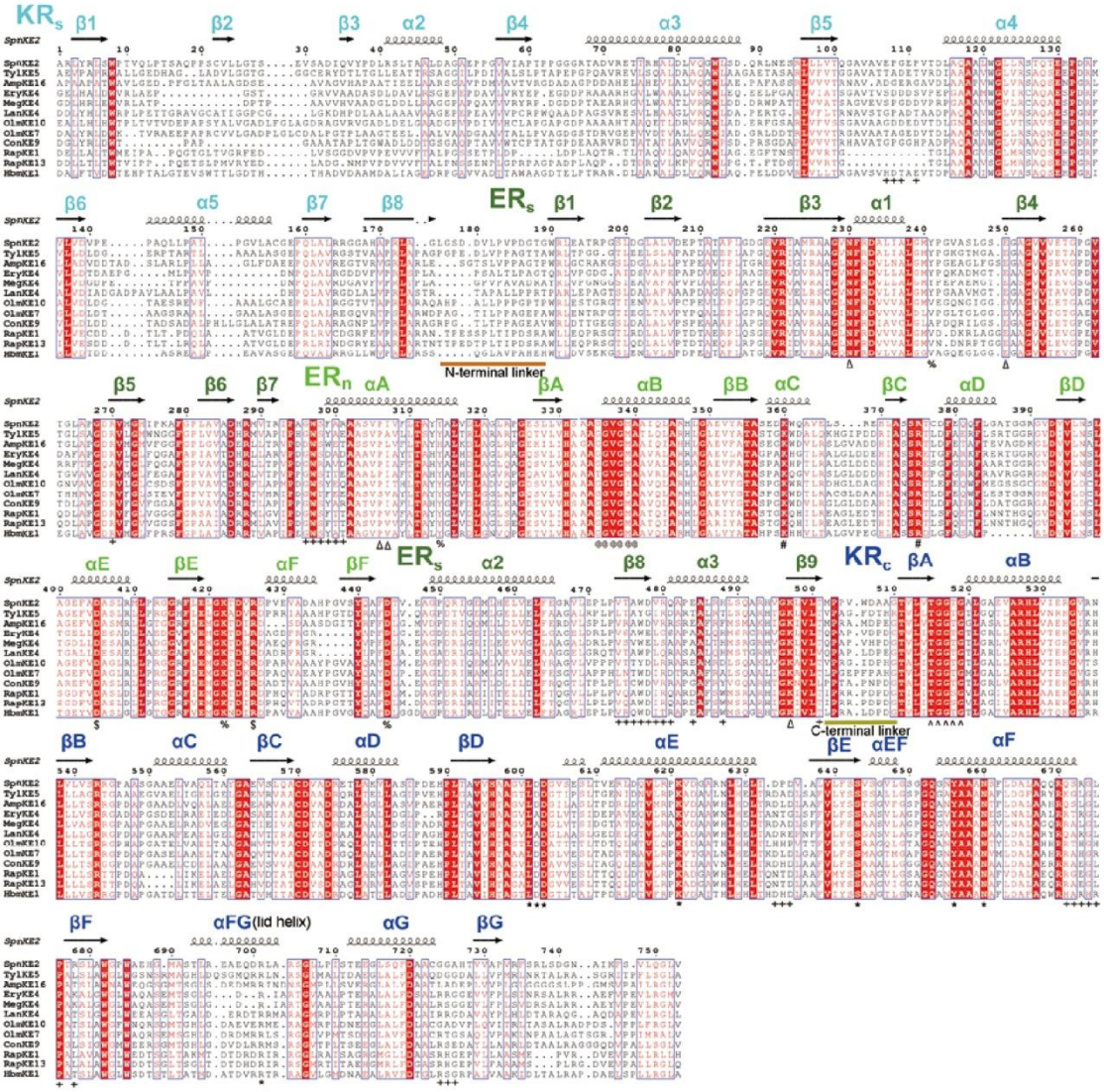


Figure 2-5. The KR/ER interface. **(a)** A stereodiagram shows interactions between several hydrophilic residues at interface between ER and the catalytic subdomain of KR. **(b)** A stereodiagram of the $2F_o$-F_c electron density map (contoured at 1.0 σ) around these residues. **(c)** One contact was observed between ER and the structural subdomain of KR.



(Figure 2-6 continued on next page.)

Figure 2-6. ER orientation and the C-terminal linker. **(a)** A stereodiagram shows that when SpnKR2 and SpnER2 are superposed on the KR and ER domains of the porcine FAS that SpnER2 is in an orientation ~180° from that observed in the Spn(KR+ER)2 crystal structure. **(b)** A stereodiagram suggests that Spn(KR+ER)2 cannot be in the FAS orientation since the C-terminal linker is not long enough to span the distance between ER and the catalytic subdomain of KR (the asterisks would need to connect). The insertion of W506 (middle of C-terminal linker) into a hydrophobic pocket of KR effectively reduces the length of the linker from 8 to 3 residues. **(c)** A stereodiagram illustrates the well-resolved electron density ($2F_o$-F_c, contoured at 1.7 σ) from the C-terminal linker of Spn(KR+ER)2. **(d)** A sequence alignment shows the conservation of the ~29-residue linker of FASs and the conservation of the 8-residue linker of polyketide synthases. GenBank accession numbers: Mouse (*Mus musculus*), NP_032014; Rat (*Rattus norvegicus*), AAA41145; Human (*Homo sapiens*), NP_004095; Goat (*Capra hircus*), ABI95140; Cow (*Bos taurus*), NP_001012687; Horse (*Equus caballus*), XP_001491342; Dog (*Canus lupus familiaris*), XP_540497; Pig (*Sus scrofa*), NP_001093400; Chicken (*Gallus gallus*), NP_990486; Frog (*Xenopus tropicalis*), XP_002937357; ConMod9 (module 9 of the concanamycin PKS from *Streptomyces neyagawensis*), AAG23265; RapMod13 (module 13 of the rapamycin PKS from *Streptomyces hygroscopicus*), CAA60462; HbmMod1 (module 1 of the herbimycin PKS from *Streptomyces hygroscopicus*), AAY28225; AmpMod16 (module 16 of the amphotericin PKS from *Streptomyces nodosus*), AAK73502; EryMod4 (module 4 of the erythromycin PKS from *Saccharopolyspora erythraea*), AAV51821; MegMod4 (module 4 of the megalomicin PKS from *Micromonospora megalomicea*), AAG13918, LanMod4 (module 4 of the lankamycin PKS from *Streptomyces avermitilis*), TylMod5 (module 5 of the tylosin PKS from *Streptomyces fradiae*), AAB66506; OlmMod10 (module 10 of the oligomycin PKS from *Streptomyces avermitilis*), BAC70603; SpnMod2 (module 2 of the spinosyn PKS from *Saccharopolyspora spinosa*) AAG23265.



(Figure 2-7 continued on next page.)

Figure 2-7. Sequence alignment of KR+ER didomains from multimodular PKSs. The secondary structure shown above the alignment is from Spn(KR+ER)2. The N- and C-terminal linkers connecting KR and ER are underlined in orange and yellow, respectively. The residues at the KR/ER interface are labeled "+". Active site residues of KR are labeled "*". The GGVGMA sequence motif of ER for NADPH binding is labeled "@". The TGGTG sequence motif of KR for NADPH binding is labeled "^". The conserved K360 and R375 that form salt bridges with the NADPH ribose phosphate are labeled "#". D405 and R426, which help orient αF, are labeled "$". The ER active site residues are labeled "%". The aligned KR+ER didomain sequences are named by their corresponding PKS and module of origin. GenBank accession numbers: SpnKE2 (module 2 of spinosyn PKS), AAG23265; TylKE5 (module 5 of tylosin PKS), AAB66506; AmpKE16 (module 16 of amphotericin PKS), AAK73502; EryKE4 (module 4 of amphotericin PKS), CAM00064; MegKE4 (module 4 of megalomicin PKS), AAG13918; LanKE4 (module 4 of lankamycin PKS), BAC76492; OlmKE10 (module 10 of oligomycin PKS), BAC70603; OlmKE7 (module 7 of oligomycin PKS), BAC70608; ConKE9 (module 9 of concanamycin PKS), AAZ94389; RapKE1 (module 1 of rapamycin PKS), CAA60460; RapKE13 (module 13 of rapamycin PKS), CAA60462; HbmKE1 (module 1 of herbimycin PKS), AAY28225.

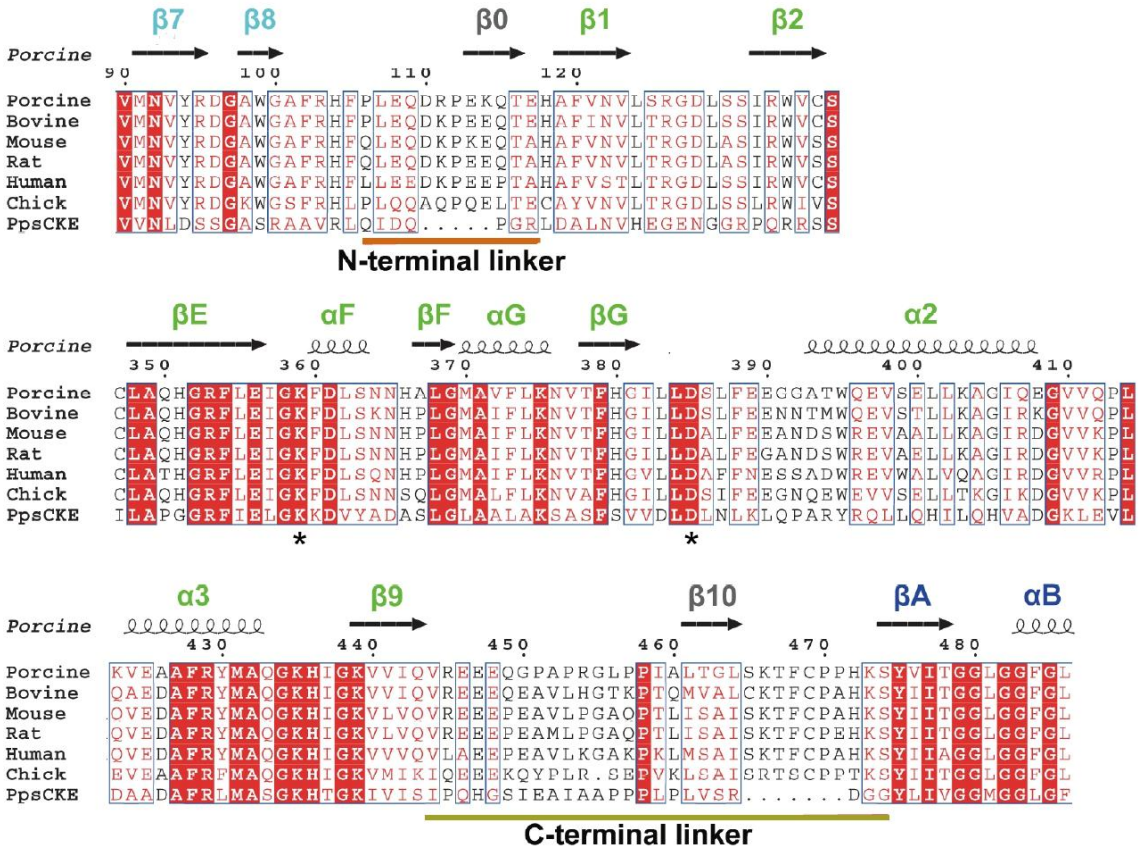


Figure 2-8. Sequence alignments of regions within ERs of mammalian FASs and PpsC. The top alignment shows the N-terminal linkers connecting KR and ER (underlined in orange). The bottom alignment shows the region mediating ER dimerization and the close sequence similarity of PpsC to the mammalian FASs (compare with the sequences surrounding βF in the multimodular PKSs in Figure 2-7). ER active site residues are labeled by asterisks. GenBank accession numbers: Porcine (*Sus scrofa*), NP_001093400; Bovine (*Bos taurus*), Q71SP7; Mouse (*Mus musculus*), P19096; Rat (*Rattus norvegicus*), P12785; Human (*Homo sapiens*), P49327; Chick (*Gallus gallus*), P12276; PpsCKE (*Mycobacterium tuberculosis*), ZP_03433948.

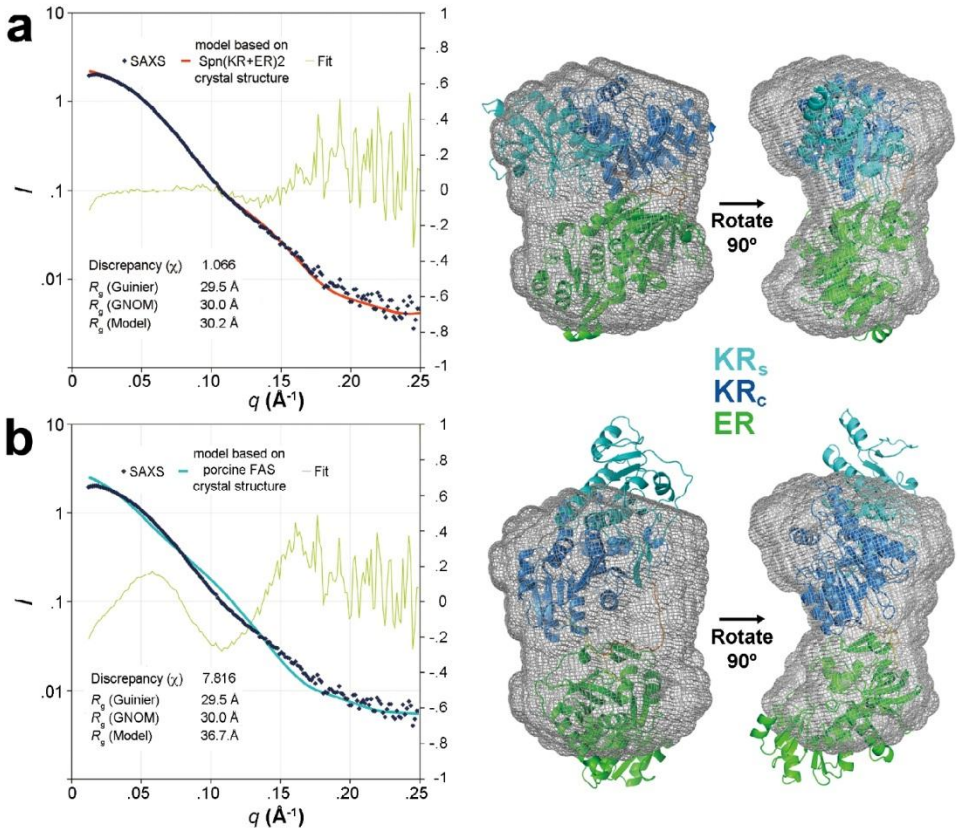


Figure 2-9. SAXS analysis of Spn(KR+ER)2. **(a)** The SAXS experimental scattering data is compared to a curve predicted by the program CRYSOL from a model of Spn(KR+ER)2 based on its crystal structure. The green line labelled "Fit" indicates the discrepancy between the SAXS data and the CRYSOL-predicted curve. The low discrepancy score indicates a good fit. An *ab initio* molecular envelope was also predicted by the program DAMMIF, and the Spn(KR+ER)2 model was fit to this envelope by the program SUPCOMB. **(b)** The curve predicted by CRYSOL from a model generated by superposing SpnKR2 and SpnER2 on the KR and ER domains of the porcine FAS is quite different from the experimental scattering data, as indicated by the high discrepancy score. This model does not fit the *ab initio* molecular envelope well. The experimental radius of gyration (R_g) determined from the Guinier plot and the program GNOM were also much closer to R_g predicted by CRYSOL for the model based on the Spn(KR+ER)2 crystal structure.

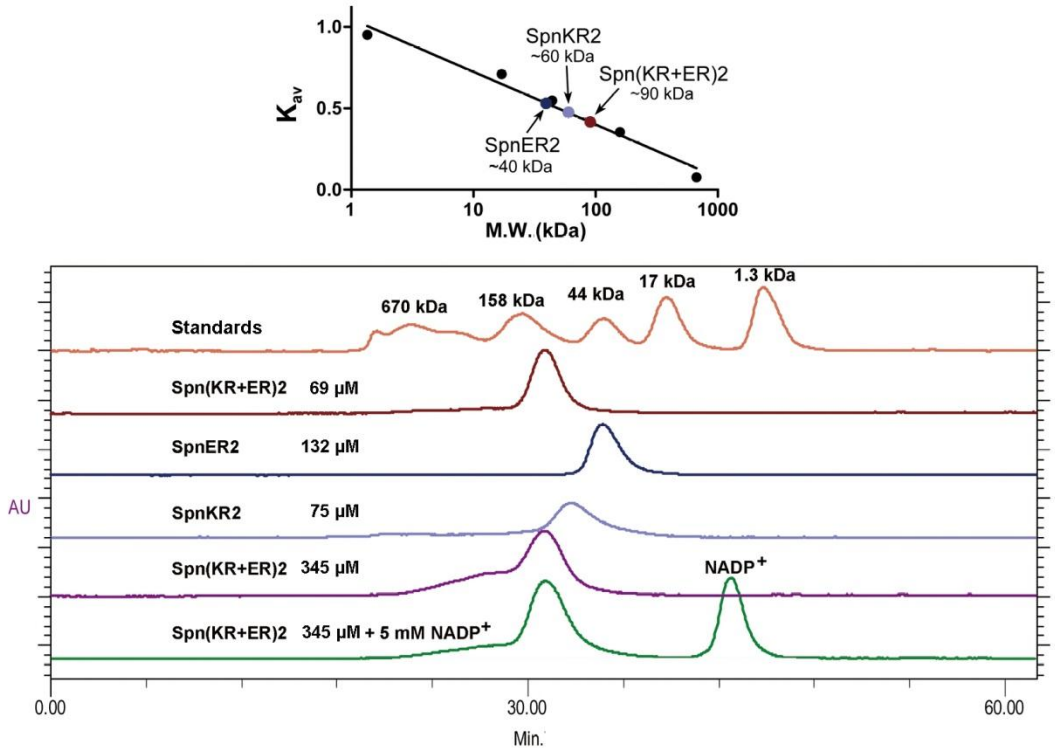
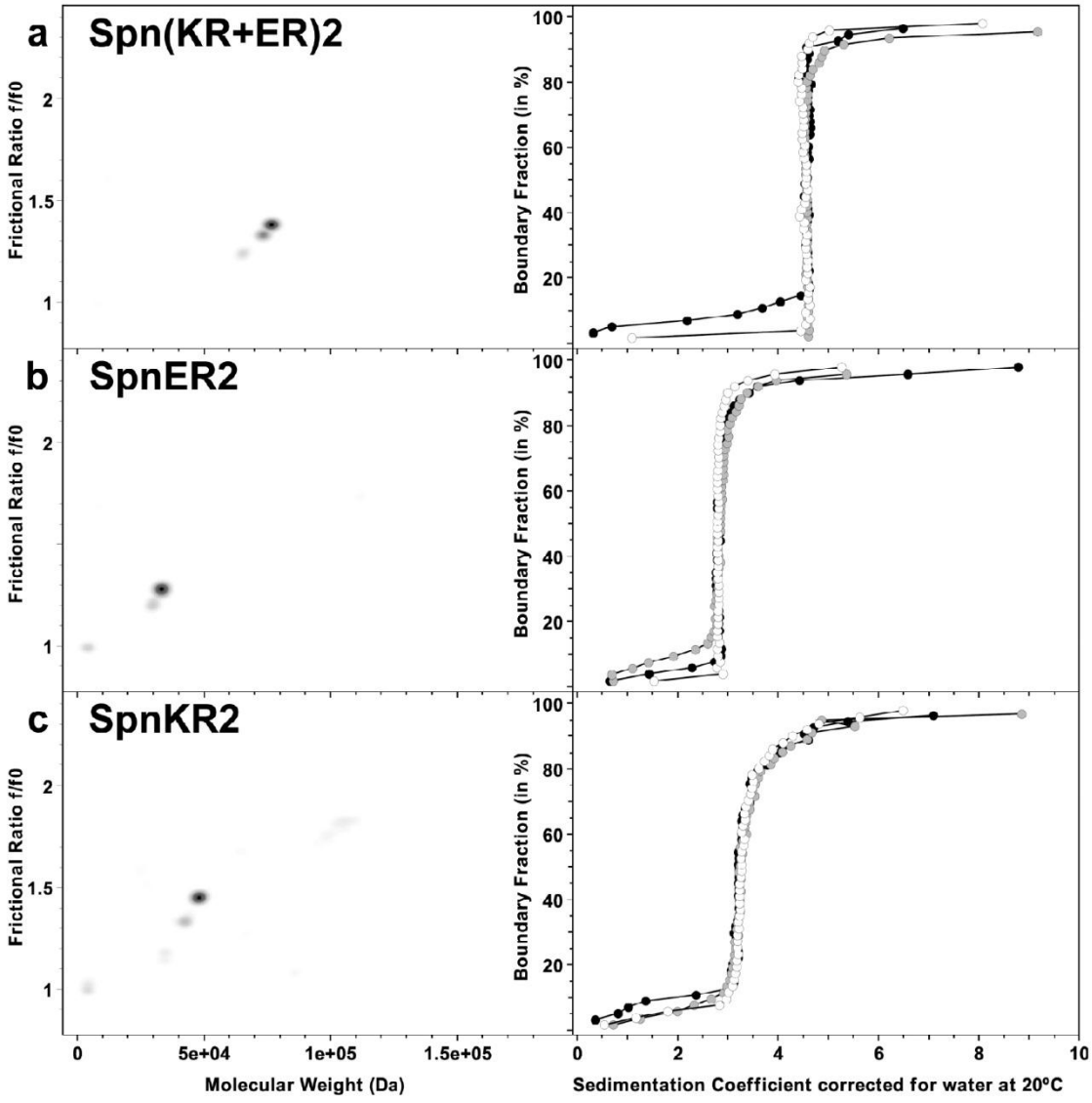


Figure 2-10. Molecular weight estimation by size-exclusion chromatography. SpnER2 migrates at ~40 kDa (expected monomer mass: 36 kDa), SpnKR2 migrates at ~60 kDa (expected monomer mass: 51 kDa), and Spn(KR+ER)2 migrates at ~90 kDa (expected monomer mass: 83 kDa) compared to standards (running buffer: 150 mM NaCl, 10 mM sodium phosphate, pH 7.4). Domain-swapping does not occur in solution since dimerization was not observed with an injection of 345 µM Spn(KR+ER)2 (purple trace); supplementing the running buffer with 0.1 mM $NADP^+$ and pre-incubating 345 µM Spn(KR+ER)2 with 5 mM $NADP^+$ also did not elicit dimerization (green trace).



d

Genetic Algorithm-Monte Carlo Analysis Results

Protein	Frictional ratio f/f0	Molecular Weight (experimental)	Molecular Weight (theoretical)
Spn(KR+ER)2	1.36 (1.22, 1.50)	7.48×10^4 (6.36×10^4, 8.60×10^4)	8.34×10^4
SpnER2	1.28 (1.19, 1.36)	3.26×10^4 (2.94×10^4, 3.59×10^4)	3.58×10^4
SpnKR2	1.46 (1.44, 1.47)	4.78×10^4 (4.70×10^4, 4.86×10^4)	5.05×10^4

(Numbers in parentheses reflect 95% confidence intervals)

(Figure 2-11 continued on next page.)

Figure 2-11. Sedimentation velocity analysis of Spn(KR+ER)2, SpnER2, and SpnKR2. **(a-c)** On the right side are the combined van Holde-Weischet integral S-value distributions of three loading concentrations measured at 280 nm (a, right, Spn(KR+ER2): black: 1.2 μM, white: 4.1 μM, and gray: 12.0 μM; b, right, SpnER2: black: 3.0 μM, white: 10.5 μM, and gray: 31 μM; c, right, SpnKR2: black: 2.5 μM, white: 7.5 μM, and gray: 23 μM). On the left side the combined genetic algorithm-Monte Carlo analysis of each protein is shown. All proteins have frictional ratios less than 1.5 and display one major species. The van Holde-Weischet analysis clearly shows no change in sedimentation coefficient and partial concentration upon change in loading concentration, indicating that in this concentration range there is no reversible association occurring and more than 80% of the sample corresponds to the monomeric form of each protein. **(d)** Molecular weights determined by the genetic algorithm analysis unambiguously show monomer only. Lower and higher molecular weight species are only minor contaminants and do not exhibit any mass action effects.

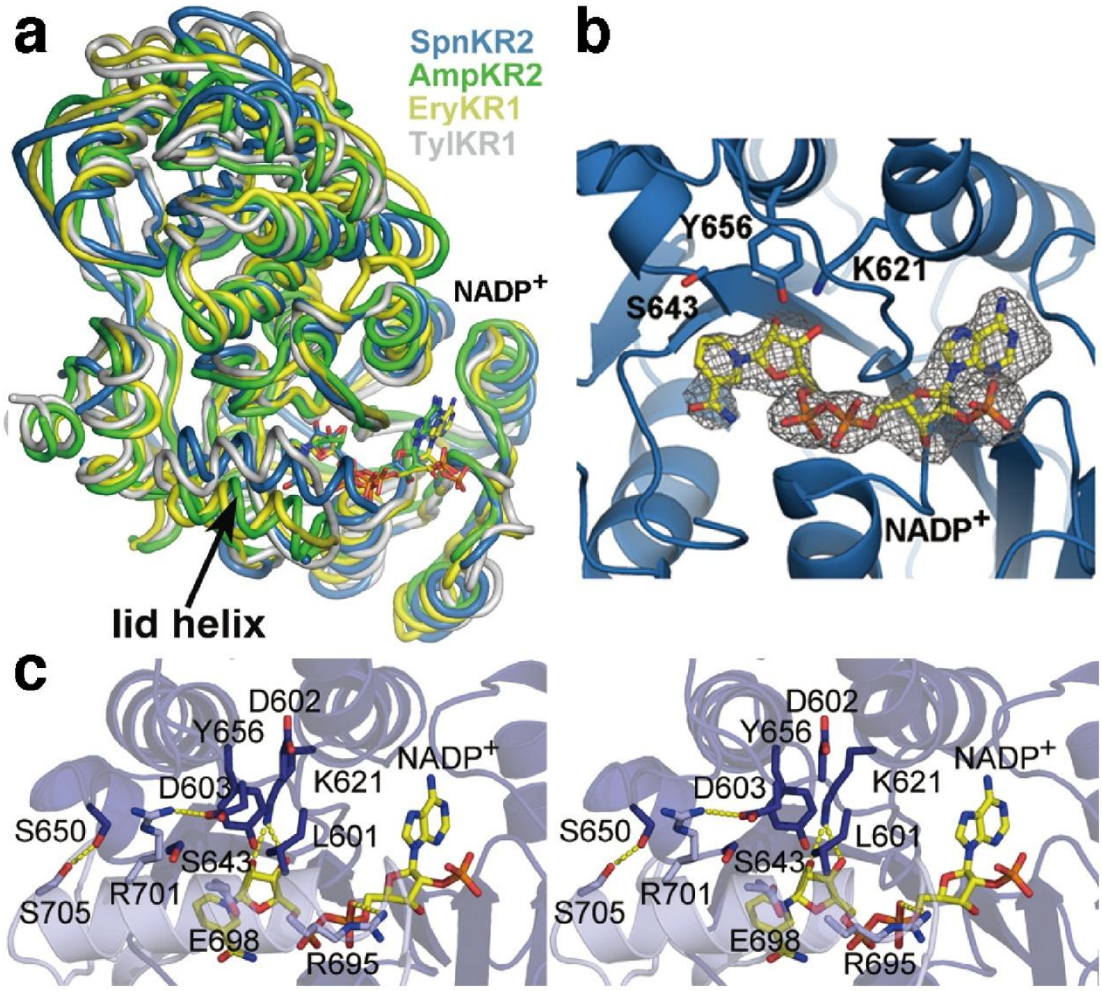


Figure 2-12. SpnKR2 close-up. **(a)** Superposition of SpnKR2 (blue), AmpKR2 (green), EryKR1 (yellow), and TylKR1 (grey). **(b)** A simulated annealing F_o-F_c omit map (contoured at 3.0 σ) shows $NADP^+$ bound at the SpnKR2 active site next to the catalytic tyrosine, lysine, serine, and asparagine. **(c)** The interactions between the lid helix (from the adjacent molecule in the asymmetric unit) and the Leu-Asp-Asp motif. Of particular note is the salt bridge formed between the nearly invariant arginine (Arg701) and the second aspartic acid (Asp603) of the Leu-Asp-Asp motif, residues known to help enforce reduction stereochemistry.

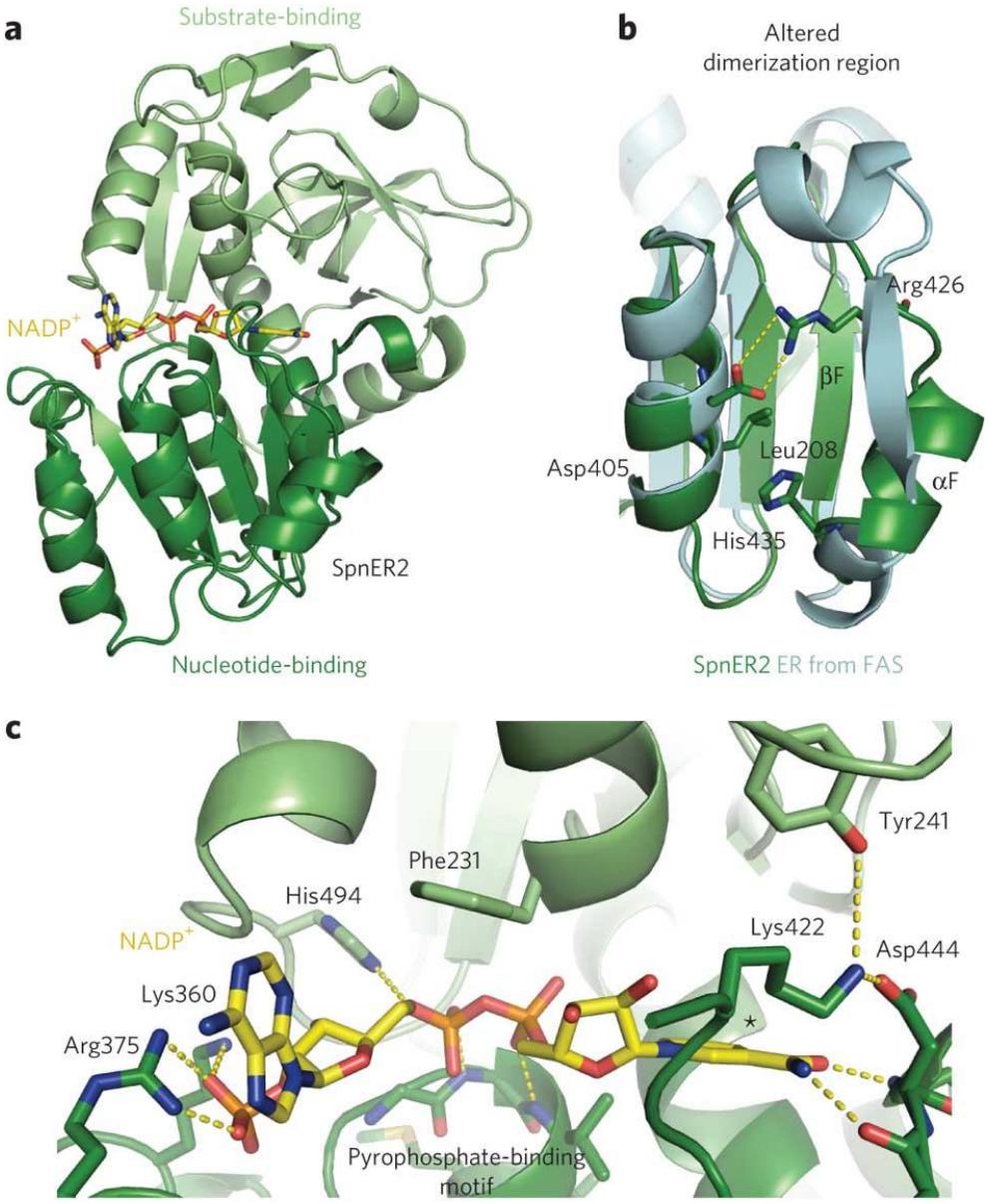


Figure 2-13. The structure of SpnER2. **(a)** SpnER2 is an MDR enzyme comprising a substrate-binding domain (residues 190–307 and 448–501) and a nucleotide-binding subdomain (residues 308–447). **(b)** A superposition of SpnER2 (green) and the ER from the porcine FAS (cyan) shows the different organization of the region surrounding βF, which usually forms the dimerization interface in MDR enzymes. **(c)** Conserved SpnER2 active site residues form specific interactions with a bound $NADP^+$ molecule. Next to where the 4-pro-R hydride of NADPH would be positioned during the reduction reaction (indicated by an asterisk) are the lysine and aspartate implicated in catalysis as well as a tyrosine known to mediate stereoselectivity.

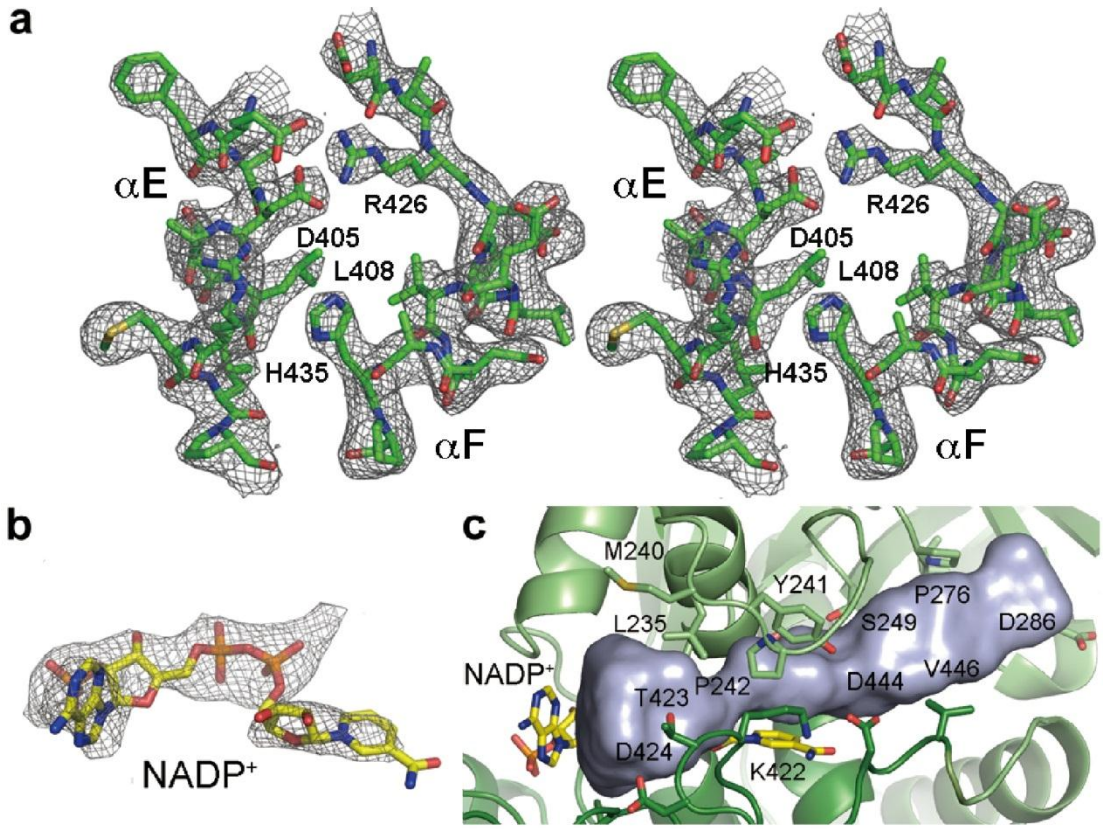


Figure 2-14. SpnER close-up. **(a)** A stereodiagram shows the $2F_o$-F_c electron density map (contoured at 1.0 σ) surrounding helices in the region usually employed by MDR enzymes for dimerization. **(b)** A simulated annealing F_o-F_c omit map (contoured at 2.5 σ) of $NADP^+$ bound at the SpnER2 active site. **(c)** The substrate binding tunnel (residues defining the tunnel are in sticks) that a polyketide intermediate would enter to be reduced by SpnER2.

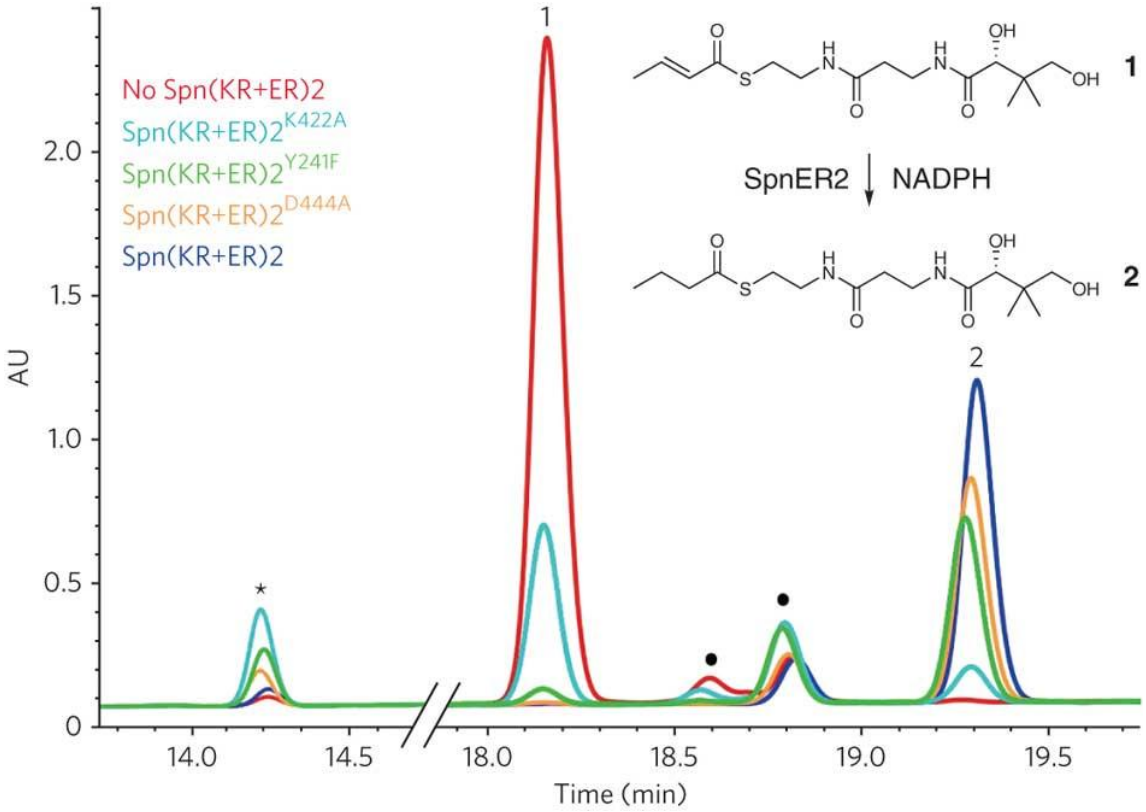


Figure 2-15. Functional analysis of SpnER2. SpnER2 reduces the α,β-unsaturated substrate analog crotonyl-pantetheine. To determine the relative contributions of SpnER2 active site residues to catalysis, three Spn(KR+ER)2 mutants were generated. Unmutated Spn(KR+ER)2 and each of the point mutants were incubated with crotonyl-pantetheine for 2 d before analysis by reversed-phase HPLC. The negative control (red) does not contain Spn(KR+ER)2. Unmutated Spn(KR+ER)2 (dark blue) completely converts crotonyl-pantetheine to butyryl-pantetheine. D444A (yellow) and Y241F (green) mutants are still active, whereas the K422A mutant (cyan) was much less so, revealing the importance of the lysine in catalysis. Each reaction contains trace impurities (filled circles). Reactions with ER mutants generate larger quantities of a byproduct (asterisk) with a mass equivalent to that of β-hydroxybutyryl-pantetheine. Details are in Figures 2-16 and Table 2-3. AU, arbitrary units.

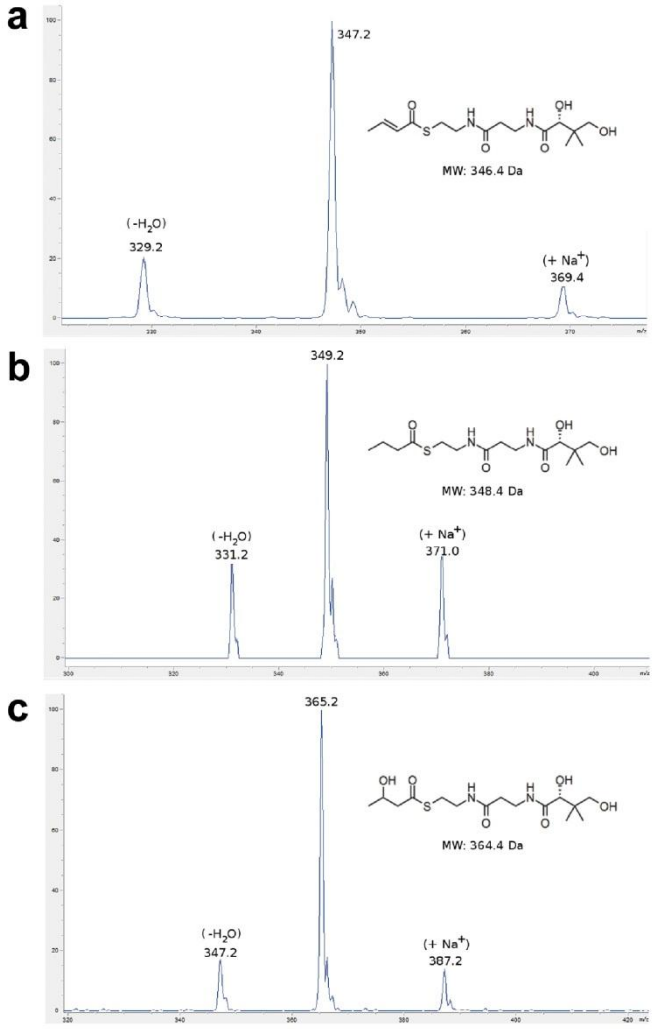


Figure 2-16. Characterization of ER reactions by mass spectrometry. **(a-c)** The crotonyl-pantetheine substrate, the butyryl-pantetheine product, and a side product that has the mass of β-hydroxybutyryl-pantetheine (labeled 1, 2, and * in Figure 2-15) were analyzed by LC/MS.

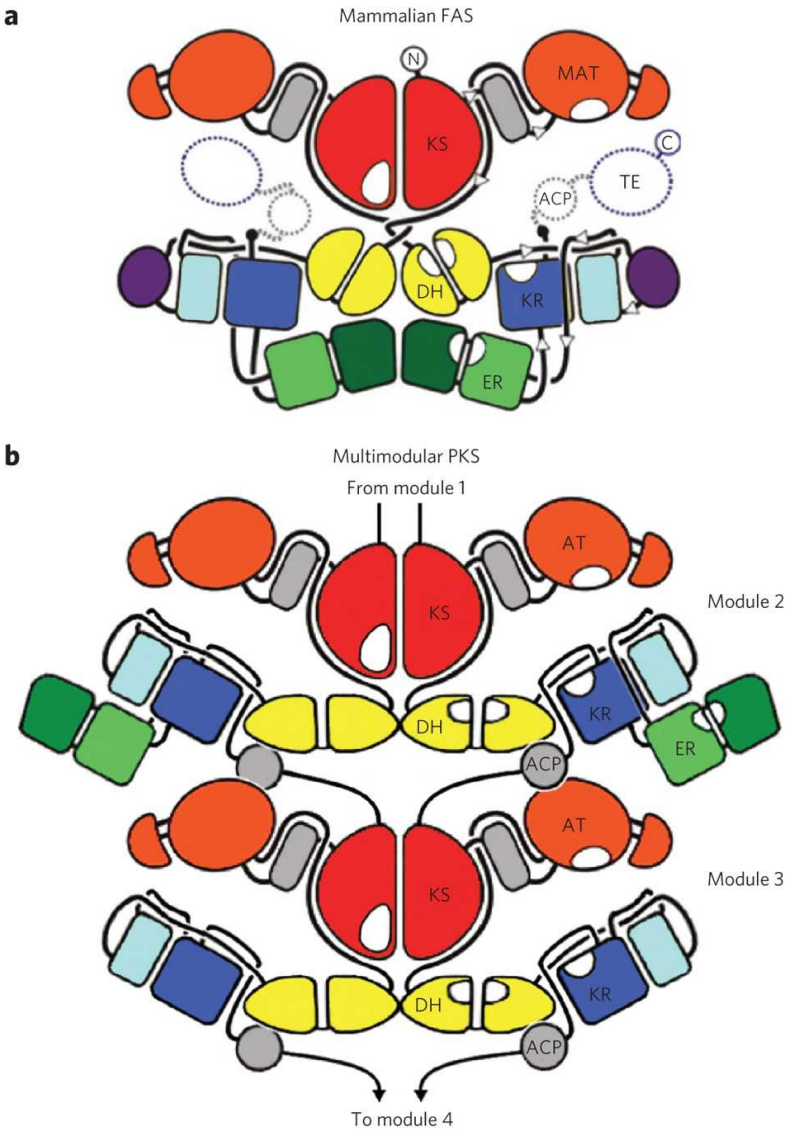


Figure 2-17. Synthase schematic. **(a)** The structure of the mammalian FAS revealed that the ketosynthase (KS), DH and ER comprise the dimer interface. **(b)** The data presented here indicate that the ERs of multimodular PKSs are monomeric and loop out from the two-fold axis. ACPs from a complete module of a multimodular PKS access all of the enzymes within that module as well as a downstream dimeric KS or dimeric thioesterase (TE). Already restrained by the short peptide that joins them to the downstream KS or TE, ACPs would be prevented from accessing their cognate enzymes by a dimeric ER. In the illustrated mycolactone multimodular PKS, the ACPs of the complete second module can access each of their cognate enzymes because ER is looped out. Artwork was reproduced with permission from N. Ban (Maier *et al.*, 2008). MAT, malonyl/acetyl-CoA transacylase. Circled N and C indicate N- and C-termini.

Data Collection				
Dataset	Native	Remote (λ1)	Inflection (λ2)	Peak (λ3)
Wavelength (Å)	0.9999	0.9645	0.9797	0.9796
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, *b*, *c* (Å)	73.1, 110.2, 202.9	73.3, 110.7, 203.4	73.4, 110.7, 203.4	73.3, 110.7, 203.4
Resolution (Å)	50-3.0(3.05-3.0)	50-3.40(3.46-3.40)	50-3.4(3.46-3.40)	50-3.4(3.46-3.40)
R_{merge}	0.10(0.54)	0.15(0.69)	0.16(0.69)	0.15(0.61)
$I / \sigma(I)$	17.9(2.8)	12.0(2.1)	12.0(2.0)	13.0(2.5)
Completeness (%)	99.4(96.4)	98.0(88.2)	97.5(84.8)	98.3(88.4)
Redundancy	5.8(5.7)	12.1(10.4)	12.0(9.8)	12.3(8.7)
Refinement				
Resolution (Å)	50-3.00			
No. reflections	31,804			
R_{work} / R_{free}	0.210/0.256			
No. atoms				
Protein	11,014			
$NADP^+$	192			
SO_4	15			
B-factors (Å^2)				
Protein	64.8			
$NADP^+$	60.4			
SO_4	53.5			
R.m.s. deviations				
Bond lengths (Å)	0.007			
Bond angles (°)	1.014			

Table 2-1. Data collection and refinement statistics.

Enoylreductase	K_m (mM)[a]	$k_{cat(R)}$ (s^{-1})[b]	$k_{cat(R)}/K_m$ ($M^{-1}s^{-1}$)
SpnER2	74 ± 14	0.0085 ± 0.0007	0.11 ± 0.02
SpnER2(D444A)	6.3 ± 0.7	0.0019 ± 0.0001	0.30 ± 0.04
SpnER2(Y241F)	24 ± 3	0.0020 ± 0.0001	0.08 ± 0.01
SpnER2(K422A)	55 ± 11	0.0010 ± 0.0001	0.018 ± 0.004

[a] K_m is defined as $(k_{-1}+k_{cat(R)}+k_{cat(H)})/k_1$. ($k_{cat(H)}$ is the turnover number of the hydration side-reaction)

[b] $k_{cat(R)}$: turnover number of the reduction reaction

Table 2-2. Steady-state kinetic parameters for reduction of crotonyl-pantetheine by SpnER2 and its point mutants. Data reported as mean ± SEM.

Enzyme	$A_{hydrated}$	$A_{reduced}$	$A_{hydrated}/A_{reduced}$
SpnER2	276533	8583817	0.03
SpnER2(D444A)	4916666	5759500	0.85
SpnER2(Y241F)	4582529	3647192	1.26
SpnER2(K422A)	5204331	749310	6.95

Reactions (20 mM substrate, two-hour incubation) were analyzed by LC/MS and the peak areas of the hydrated (365.2 Da) and reduced (349.2 Da) products were obtained form the extracted ion chromatograms respectively.

$A_{hydrated}$: peak area of hydrated product

$A_{reduced}$: peak area of reduced product

$A_{hydrated}/A_{reduced}$: peak area ratio of hydrated and reduced products

Table 2-3. Comparison of hydrated and reduced products generated by SpnER2 and its point mutants.

Chapter 3. Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase

ABSTRACT

RifDH10, the dehydratase domain from the terminal module of the rifamycin polyketide synthase, catalyzes the stereospecific syn dehydration of the model substrate (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-RifACP10, resulting in the exclusive formation of (*E*)-2-methyl-2-pentenoyl-Rif-ACP10. RifDH10 does not dehydrate any of the other three diastereomeric, RifACP10-bound, diketide thioester substrates. On the other hand, when EryACP6, from the sixth module of the erythromycin polyketide synthase, is substituted for RifACP10, RifDH10 stereospecifically dehydrates only (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-EryACP6 to give exclusively (*E*)-2-methyl-2-pentenoyl-EryACP6, with no detectable dehydration of any of the other three diastereomeric, EryACP6-bound, diketides. An identical alteration in substrate diastereospecificity was observed for the corresponding N-acetylcysteamine or pantetheine thioester analogues, regardless of acyl chain length or substitution pattern. Incubation of (2*R*S)-2-methyl-3-ketopentanoyl-RifACP10 with the didomain reductase-dehydratase RifKR10-RifDH10 yielded (*E*)-2-methyl-2-pentenoyl-RifACP10, the expected product of syn dehydration of (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-RifACP10, while incubation with the corresponding EryACP6-bound substrate, (2*RS*)-2-methyl-3-ketopentanoyl-EryACP6, gave only the reduction product (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-EryACP6 with no detectable

Chapter 3 was adapted from the following published work: Gay D, You YO, Keatinge-Clay AT, and Cane DE. (2013) Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. *Biochemistry* **52**, 8916−8928. Each author contributed to the experimental design and manuscript preparation.

dehydration. These results establish the intrinsic syn dehydration stereochemistry and substrate diastereoselectivity of RifDH10 and highlight the critical role of the natural RifACP10 domain in chaperoning the proper recognition and processing of the natural ACP-bound undecaketide substrate. The 1.82 Å-resolution structure of RifDH10 reveals the atomic-resolution details of the active site and allows modeling of the syn dehydration of the (2*S*,3*S*)-2-methyl-3-hydroxyacyl-RifACP10 substrate. These results suggest that generation of the characteristic *cis* double bond of the rifamycins occurs after formation of the full-length RifACP10-bound acyclic *trans*-unsaturated undecaketide intermediate, most likely during the subsequent macrolactamization catalyzed by the amide synthase RifF.

INTRODUCTION

Polyketides make up an enormously diverse group of microbial natural products, many of which are important antibacterial, antifungal, immunosuppressive, and anticancer agents. The multimodular polyketide synthases (PKSs) that assemble complex polyketides are structurally and mechanistically related to the metazoan fatty acid synthases (FASs) (Smith & Tsai, 2007; Sherman & Smith, 2006; Cane, 2010). While FASs condense two-carbon acetate units and reduce newly formed β-ketoacyl groups in an iterative fashion, multimodular PKSs exploit an assembly line logic in which 5−25 modules (sets of enzymes similar to those in an FAS) each perform a decarboxylative chain-elongation reaction that adds successive malonyl, methylmalonyl, or ethylmalonyl building blocks to the growing polyketide and processing reactions that control the

oxidation level and stereochemistry of the substituents attached to the α- and β-positions of the newly extended chain.

Type I PKS modules contain at least three independently folded catalytic domains: (1) an acyltransferase (AT) that selects a unique malonyl-, methylmalonyl-, or ethylmalonyl-CoA chain extender from the cellular acyl-CoA pool, (2) an acyl carrier protein (ACP) that acquires the extender unit from the AT through a flexible, 18 Å phosphopantetheine arm, and (3) a β-ketoacyl-ACP synthase [ketosynthase (KS)] that catalyzes the decarboxylative chain-elongation reaction between a growing polyketide transferred to the KS by the upstream module and the ACP-tethered α-carboxyacyl extender unit (the closely related *trans*-AT PKS modules lack an integrated AT domain) (Piel, 2010). Most modules also contain one or more processing domains, such as a ketoreductase (KR) that stereoselectively catalyzes the NADPH-dependent reduction of the β-keto group and can stereospecifically set the configuration of a neighboring α-substituent, a dehydratase (DH) that eliminates the resulting β-hydroxyl group as well as an α-proton to form an α,β-enoyl double bond, and an enoylreductase (ER) that catalyzes the stereoselective, NADPH-dependent reduction of the double bond. During polyketide assembly, growing chains remain covalently bound through a thioester linkage either to the phosphopantetheinyl thiol of an ACP or, transiently, to the active site cysteine residue of the KS domain, with the relatively mobile ACP domains shuttling the attached substrates among processing domains and to downstream modules. In most PKSs, the release of the fully elongated and processed polyketide is catalyzed by a dedicated

thioesterase (TE) domain located at the C-terminal end of the ultimate module in the synthase, resulting in either macrolactonization or hydrolysis.

The vast majority of the more than 2000 known complex polyketides contain one or more double bonds, most of which have *trans* (*E*) geometry. A minor but nonetheless significant fraction of complex polyketides also harbor one or more *cis* (*Z*) double bonds (**Figure 3-1**). How these *cis* double bonds are installed in polyketides such as the protein phosphatase inhibitors fostriecin (Lewy *et al.*, 2002; Kong *et al.*, 2013) and phoslactomycin (Palaniappan *et al.*, 2008), the anti-angiogenic agent borrelidin (Olano *et al.*, 2004), the microtubule stabilizer epothilone (Tang *et al.*, 2000), the microtubule polymerization inhibitor curacin A, (Akey *et al.*, 2010) and the rifamycin family of antibacterials (Schupp *et al.*, 1998; August *et al.*, 1998; Tang *et al.*, 1998) remains largely unknown. Reynolds and co-workers have recently established that Plm1, the first module of the phoslactomycin PKS, most likely produces *cis*-3-cyclohexylpropenoate and that the PlmKR1 domain generates a (3*S*)-hydroxyacyl thioester intermediate that serves as the substrate for the PlmDH1 domain (Alhamadsheh *et al.*, 2007; Bonnett *et al.*, 2013).

The ansamycin antibiotic rifamycin B (**Figure 3-2**) is produced by the actinomycete *Amycolatopsis mediterranei*. The semisynthetic derivatives rifapentine, rifampicin, and rifabutin inhibit the RNA polymerase of mycobacteria and are clinically prescribed to tuberculosis patients, with rifampicin listed by the World Health Organization as one of the five essential antituberculosis drugs (World Health Organization, 2010). Rifamycins are distinguished from the more common macrolide and polyene polyketides by several unique structural features, including an

aminonaphthoquinone, a macrolactam ring, and a trisubstituted *cis* double bond (Hertweck, 2009). The rifamycin biosynthetic gene cluster harbors a 10-module PKS that extends a 3-amino-5-hydroxybenzoate (AHBA) starter unit with two acetates and eight propionates (Schupp *et al.*, 1998; August *et al.*, 1998; Tang *et al.*, 1998). The rifamycin PKS lacks the typical TE domain fused to terminal module 10. Instead, release of the fully processed, RifACP10-bound undecaketide is catalyzed by the separately encoded amide synthase, RifF. Although two research teams have independently reported that mutation or deletion of *rifF* leads to release of acyclic 2-methyl-2-enoyl undecaketides, these investigators reached contradictory conclusions regarding the geometry of the 15,16 double bond (rifamycin numbering) of this abortive product (Stratmann *et al.*, 1999; Yu *et al.*, 1999; Floss *et al.*, 1999). The geometry of the double bond in the fully processed RifACP10-undecaketide intermediate produced by the rifamycin PKS, which serves as the actual substrate for cyclization by amide synthase RifF, therefore remains unsettled, nor is it known whether RifDH10 itself or the cyclase RifF sets the characteristic *cis* double bond geometry of the rifamycins (**Figure 3-2**).

Neither sequence nor structural alignments have revealed any features that correlate DH protein sequence with the *trans* or *cis* geometry of the double bond in the resultant product (Keatinge-Clay, 2008). For example, the active site of EryDH4, which catalyzes the formation of a trisubstituted *trans* (*E*) double bond (Donadio *et al.*, 1993; Valenzano *et al.*, 2010) is structurally equivalent to that established for the DH domains of curacin PKS subunits CurF and CurH, one of which is thought to generate a *cis* double bond (Akey *et al.*, 2010). DH-containing modules that generate *trans* double bonds are

often paired with KR domains that produce (3*R*)-hydroxyacyl products. Such KR domains are readily recognized by their characteristic amino acid sequence fingerprints (Reid *et al.*, 2003; Caffrey, 2003; Keatinge-Clay, 2007; Castonguay *et al.*, 2008). On the other hand, there is only limited circumstantial evidence correlating KR domains that generate (3*S*)-hydroxyacyl intermediates with the generation of *cis* double bonds by their paired DH domains (Keatinge-Clay, 2008). For example, KR domains may provide (3*S*)-hydroxyacyl substrates for the DH domains of modules 2 and 3 of the fostriecin PKS that generate *cis* double bond diketide and triketide intermediates (Kong *et al.*, 2013), as well as for the DH domains of modules 1 and 2 of the phoslactomycin PKS (Alhamadsheh *et al.*, 2007; Bonnett *et al.*, 2013; Palaniappen *et al.*, 2003), module 4 of the epothilone PKS (Tang *et al.*, 2004), the CurG module of the curacin PKS (Akey *et al.*, 2010), and module 10 of the rifamycin PKS (Schupp *et al.*, 1998; August *et al.*, 1998; Tang *et al.*, 1998). Whether modules that contain both an A-type KR and a DH actually generate *cis* double bonds, however, is still largely untested, except in the case of phoslactomycin module 1 mentioned above (Alhamadsheh *et al.*, 2007; Bonnett *et al.*, 2013). As a further complication, the formation of some *cis* double bonds can involve post-PKS transformations such as dehydration (Palaniappan *et al.*, 2008) or double-bond isomerization (Vergnolle *et al.*, 2011).

We have recently established that RifKR10, the KR domain from module 10 of the rifamycin PKS, and RifKR7 from module 7 of the same PKS each mediate the stereospecific epimerization and/or reduction of (2*R*)-2-methyl-3-ketoacyl-ACP substrates to give exclusively (2*S*,3*S*)-2-methyl-3-hydroxyacyl-ACP products (You *et al.*,

2013). The RifKR10-generated product must serve as the substrate for the paired dehydratase domain of module 10, RifDH10. Although the 3*S* configuration of this 3-hydroxyacyl substrate might superficially suggest the formation of a *cis*-2-methyl-2-enoyl-ACP product, this geometry cannot be generated by syn dehydration of a (2*S*,3*S*)-2-methyl-3-hydroxyacyl intermediate. In fact, all DH domains that have been stereochemically characterized to date, including EryDH4 from the erythromycin synthase (Valenzano *et al.*, 2010), NanDH2 from the nanchangmycin synthase (Guo *et al.*, 2010), and TylDH2 from the tylactone synthase (Castonguay *et al.*, 2008), display strict specificity for the diastereomeric (2*R*,3*R*)-2-methyl-3-hydroxyacyl-ACP substrates and catalyze completely stereospecific syn dehydration to give the corresponding *trans*-(*E*)-2-methyl-2-enoyl-ACP products. In light of not only its unusual substrate stereochemistry but also the uncertainty in the actual *cis* or *trans* geometry of its natural undecaketide product, the RifDH10-catalyzed dehydration reaction presents an intriguing and important mechanistic, stereochemical, and biosynthetic puzzle.

We now report the expression and purification of the recombinant RifDH10 domain, the determination of its protein structure to 1.82 Å resolution, and the demonstration that RifDH10 catalyzes the diastereospecific syn dehydration of (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-RifACP10 to give exclusively the corresponding *trans*-(*E*)-2-methyl-2-pentenoyl thioester product. Unexpectedly, the natural diastereospecificity of RifDH10 is reversed when the dehydratase is incubated with thioester substrates that are tethered to the noncognate ACP domain EryACP6 or to the corresponding pantetheinyl or N-acetylcysteamine (NAC) analogues, resulting in the anomalous syn dehydration of

only the diastereomeric (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl thioesters to give the corresponding (*E*)-2-methyl-2-pentenoyl thioester products.

RESULTS

Expression and purification of RifDH10 and RifACP10

To investigate the mechanism and structure of RifDH10, we used a synthetic gene with codons optimized for expression in *E. coli* to express the RifDH10 domain as a discrete recombinant protein corresponding to the region from A2569 to G2872 of RifE (EMBL AAC01714.1; UniProt entry O54593; residues herein numbered 1−304), based on alignment with the previously described stand-alone EryDH4 domain (Keatinge-Clay, 2008; Valenzano *et al.*, 2010) and flanked by the deduced C-terminal and N-terminal boundaries, respectively, of the RifAT10 and RifKR10 domains (August *et al.*, 1998; Tang *et al.*, 1998) (**Figures 3-3** and **3-4**). After ligation of the synthetic gene into pET28a and transformation into *E. coli* BL21(DE3), the expressed recombinant RifDH10 carrying an N-terminal His_6 tag was purified by Ni(II) affinity chromatography and analyzed for purity and Mr by SDS−PAGE (**Figure 3-5**). Attempted expression of recombinant RifACP10 corresponding to the region from L3324 to L3409 of RifE, using a synthetic RifACP10 gene with codons optimized for expression in *E. coli* (**Figure 3-6**), gave only insoluble inclusion bodies. Instead, apo-RifACP10-NusA (hereafter simply termed apo-RifACP10) could be obtained as a soluble fusion protein carrying an N-terminal His_6 tag and a C-terminal NusA (De Marco, 2004). The resultant protein was readily purified by Ni(II) affinity chromatography (**Figure 3-5**).

Stereochemistry of RifDH10-catalyzed dehydration of 2-methyl-3-hydroxyacyl-RifACP10

We have recently reported that (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-RifACP10 (1a) can be prepared by stereospecific reduction of (2*R*S)-2-methyl-3-ketopentanoyl-RifACP10 with either RifKR7 or RifKR10 in the presence of NADPH (**Figure 3-7a** and **3-9**) (You *et al.*, 2013). The requisite RifACP10-bound 3-keto acylthioester was generated *in situ* by chemoenzymatic incubation of (2*R*S)-2-methyl-3-ketopentanoyl-CoA with apo-RifACP10 and the phosphopantetheinyl transferase Sfp (You *et al.*, 2013). Incubation of (2*S*,3*S*)-1a with RifDH10 gave as the exclusive dehydration product (*E*)-2-methyl-2-pentenoyl-Rif-ACP10 (2), as established by enzymatic hydrolysis with PicTE, the thioesterase domain from the picromycin synthase, followed by GC−MS analysis and direct comparison of the derived (*E*)-2-methyl-2-pentenoate (3) with an authentic sample of 3, as well as with the isomeric (*Z*)-2-methyl-2-pentenoic acid (**Figures 3-7a** and **3-8**) (Valenzano *et al.*, 2010). Omission of RifDH10 resulted in exclusive formation of (2*S*,3*S*)-1a, as confirmed by PicTE-catalyzed hydrolysis and chiral GC−MS analysis of the derived methyl ester (2*S*,3*S*)-4a, which was identical in both retention time and mass spectrum by direct comparison with authentic 4a, as previously observed (**Figure 3-9**) (You *et al.*, 2013). By contrast, incubations of RifDH10 under identical conditions with any of the chemoenzymatically prepared diastereomeric substrates, (2*R*,3*R*)-1b (TylKR1), (2*R*,3*S*)-1c (EryKR6), or (2*S*,3*R*)-1d (EryKR1) (Castonguay *et al.*, 2008; Valenzano *et al.*, 2009; Castonguay *et al.*, 2007) resulted at most in generation of only very minor quantities (<5−10%) of 3, most likely because of the previously reported

(You *et al.*, 2013) generation of minor amounts of contaminating (2*S*,3*S*)-1a by TylKR1 and EryKR1 upon incubation with (2*R*S)-2-methyl-3-ketopentanoyl-RifACP10 (**Figures 3-10**, **3-11**, and **3-12**).

We have previously established that RifKR10 has a stereospecificity identical to that of RifKR7, reducing (2*R*S)-2-methyl-3-ketopentanoyl-RifACP10 to the corresponding (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-RifACP10 (1a) (You *et al.*, 2013). To confirm that the observed stereochemistry of the coupled RifDH10-catalyzed dehydration reaction is not influenced by the origin of the 2*S*,3*S*-specific ketoreductase domain, we prepared a RifDH10-RifKR10 fusion protein. Incubation of Rif[DH10][KR10] with (2*R*S)-2-methyl-3-ketopentanoyl-Rif-ACP10 and NADPH gave (*E*)-2-methyl-2-pentenoyl-RifACP10 (2), as established by GC−MS analysis of the derived (*E*)-2-methyl-2-pentenoic acid (3) (**Figures 3-7c** and **3-13**). Analysis of the methyl ester derived from the reduced diketide intermediate 1a confirmed the expected exclusive formation of (2*S*,3*S*)-4a.

syn stereochemistry of RifDH10-catalyzed hydration of (*E*)-2-methyl-2-pentenoyl-RifACP10

In a complementary experiment, we also investigated the reverse RifDH10-catalyzed hydration of (*E*)-2-methyl-2-pentenoyl-RifACP10 (2). Incubation of chemoenzymatically prepared (*E*)-2-methyl-2-pentenoyl-RifACP10 (2) resulted in stereospecific syn hydration to give exclusively (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-

RifACP10 (1a), as confirmed by PicTE hydrolysis, methylation, and chiral GC−MS detection of methyl (2*S*,3*S*)-4a (**Figures 3-7b** and **3-14**).

Stereochemistry of RifDH10-catalyzed dehydration of 2-methyl-3-hydroxyacyl thioester analogues

Although these results are all congruent with the previously established 2*S*,3*S* stereospecificity of RifKR10 (You *et al.*, 2013), which generates the natural substrate for RifDH10, the observed syn dehydration of a (2*S*,3*S*)-2-methyl-3-hydroxyacyl-ACP substrate by RifDH10 to give the corresponding (*E*)-2-methyl-2-enoylacyl-ACP product is biochemically unprecedented. Thus, all other mechanistically characterized DH domains catalyze syn dehydration of only the diastereomeric (2*R*,3*R*)-2-methyl-3-hydroxyacyl-ACP substrates to give (*E*)-2-methyl-2-enoyl-ACP products (Valenzano *et al.*, 2010; Guo *et al.*, 2010), while a (3*S*)-hydroxyacyl intermediate has been implicated in the formation of a *cis*-enoyl thioester in the biosynthesis of phoslactomycin (although the overall stereochemistry of the latter dehydration has not been determined) (Bonnett *et al.*, 2013). To explore any influence of RifACP10 on the overall stereochemistry of the RifDH10-catalyzed reaction, we therefore also examined the stereospecificity of RifDH10-catalyzed dehydration of diketide thioester substrates that are covalently tethered to alternative thioester analogues of RifACP10. To our surprise, we found that replacement of RifACP10 with the heterologous EryACP6 domain, obtained from module 6 of the erythromycin PKS, resulted in an unexpected complete alteration of the diastereospecificity of RifDH10, without any effect on either the intrinsic syn

stereospecificity of the dehydration reaction or the *trans* geometry of the resulting unsaturated product (**Figure 3-15**). Thus, although RifDH10 was unable to dehydrate chemoenzymatically prepared (2*S*,3*S*)-2-methyl-3-hydroxyacyl-EryACP6 (5a), as established by the TLC−phosphorimaging assay (**Figure 3-16**), RifDH10 did catalyze the syn dehydration of the diastereomeric anti-(2*R*,3*R*)-2-methyl-3-hydroxyacyl-EryACP6 (5b), which could be generated *in situ* by coupled incubation of the Ery[KS6][AT6] didomain from module 6 of the erythromycin PKS, EryACP6, and TylKR1 (from module 1 of the tylactone PKS) with propionyl-SNAC, methylmalonyl-CoA, and NADPH, as previously described (Valenzano *et al.*, 2010; Castonguay *et al.*, 2008; Valenzano *et al.*, 2009; Castonguay *et al.*, 2007). The reaction gave exclusively (*E*)-2-methyl-2-pentenoyl-EryACP6 (6), whose structure and stereochemistry were established by hydrolysis with 0.5 M NaOH for 20 min at 65 °C and GC− MS analysis of the derived methyl ester 3 (**Figures 3-15a** and **3-17**). Similarly, an analogous incubation in which (2*S*,3*R*)-2-methyl-3-hydroxypentanoyl-SNAC was substituted for propionyl-SNAC gave exclusively (*E*)-(4R,5R)-2,4-dimethyl-5-hydroxy-2-heptenoyl-ACP6, as established by GC−MS analysis of the derived methyl ester (**Figure 3-18**). On the other hand, RifDH10 did not dehydrate either of the diastereomeric EryACP6-bound syn thioesters, (2*R*,3*S*)-5c or (2*S*,3*R*)-5d (**Figure 3-15a** and **3-16**). RifDH10 also catalyzed the reverse syn hydration of (*E*)-2-methyl-2-pentenoyl-EryACP6 (6), chemoenzymatically prepared *in situ*, as previously described (Valenzano *et al.*, 2010) to give methyl (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-EryACP6 (5b), whose structure and stereochemistry were established by PicTE catalyzed hydrolysis and methylation, followed by chiral GC− MS analysis and direct comparison

of the derived methyl ester (2*R*,3*R*)-4b with authentic standards of each of the four diastereomeric diketide methyl esters (**Figure 3-15b** and **3-19**).

A parallel series of incubations of RifDH10 with the corresponding pantetheine and NAC thioesters also resulted in syn dehydration and hydration of only the anomalous (2*R*,3*R*)-2-methyl-3-hydroxyacyl thioester diastereomers and their corresponding (*E*)-2-methyl-2-enoylacyl thioesters (**Figure 3-15c**, **3-20**, **3-21**, **3-22**, **3-23**, **3-24**, **3-25**, and **3-26**). Thus, incubation of RifDH10 with (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-*S*-pantetheine (7b), prepared by diastereospecific reduction of (2*R*S)-2-methyl-3-ketopentanoyl-*S*-pantetheine with TylKR1 (Keatinge-Clay, 2007; Piasecki *et al.*, 2011; Siskos *et al.*, 2005) and a coupled enzymatic NADPH regenerating system (Piasecki *et al.*, 2011), gave exclusively (*E*)-2-methyl-2-pentenoyl-*S*-pantetheinate (8), as established by HPLC analysis and direct comparison with authentic synthetic standards (**Figures 3-15c**, **3-22**, and **3-23**). In agreement with this result, incubation of RifDH10 with (*E*)-2-methyl-2-butenoyl-*S*-pantetheine (11) yielded exclusively the syn hydration product (2*R*,3*R*)-2-methyl-3-hydroxybutanoyl-Spantetheine (12b) (**Figures 3-15d**, **3-25**, and **3-27**). The isomeric (*Z*)-2-methyl-2-butenoyl-*S*-pantetheine (13) did not undergo RifDH10-catalyzed hydration. RifDH10 did not dehydrate diastereomeric diketide S-pantetheinate (2*S*,3*S*)-7a, (2*R*,3*S*)-7c, or (2*R*,3*S*)-2-methyl-3-hydroxypentanoyl-*S*-pantetheine (7d). Analogous results were also obtained using the corresponding NAC thioester analogues. Thus, RifDH10 catalyzed the syn dehydration of only (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-*S*-NAC (9b) to give the corresponding (*E*)-2-methyl-2-pentenoyl-*S*-NAC (10) (**Figures 3-15c** and **3-22**). We also examined the effects of acyl chain length

and substitution pattern, observing that RifDH10 catalyzed the reversible dehydration of (3*R*)-hydroxybutanoyl-*S*-pantetheine (**Figures 3-24** and **3-27**) and the hydration of (*E*,*E*)-2,4-hexadienoyl-Spantetheine (**Figure 3-26**).

Incubation of the Rif[DH10][KR10] fusion protein with chemoenzymatically prepared (2*R*S)-2-methyl-3-ketopentanoyl-EryACP6 gave only the corresponding product of reduction by RifKR10, (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-EryACP6 (5a), with no detectable dehydration product, in contrast to the processing of the RifACP10-bound substrate, which resulted in both reduction and dehydration to give the corresponding (*E*)-2-methyl-3-pentenoyl-RifACP10 product, as described above. In a like manner, although Rif[DH1]-[KR10] produced alcohol (2*S*,3*S*)-5a by diastereospecific reduction of (2*R*)-2-methyl-3-ketopentanoyl-EryACP6, generated *in situ* by decarboxylative condensation of propionyl-*S*-NAC and methylmalonyl-CoA catalyzed by Ery[KS6][AT6] plus EryACP6, none of the corresponding dehydration product could be detected (**Figures 3-7c** and **3-28**).

Structure of RifDH10

To improve our understanding of the protein structural basis for the dehydration reaction catalyzed by RifDH10, its structure was determined to 1.82 Å resolution (**Figure 3-29** and **Table 3-1**). RifDH10 is structurally highly homologous to the other previously determined DH domains, including EryDH4 and the four DH domains from CurF, CurH, CurJ, and CurK of the curacin PKS [PDB entries 3EL6, 3KG6, 3KG7, 3KG8, and 3KG9, respectively, and rootmean-square deviations (rmsd) of 1.2, 1.8, 2.0, 2.3, and 1.7 Å,

respectively] (Akey *et al.*, 2010; Keatinge-Clay, 2008). RifDH10 possesses a characteristic double-hotdog fold and dimerizes through an interface largely created by its ~25 N-terminal residues (**Figure 3-29a**). The catalytic dyad, comprised of a histidine (H50) and an aspartic acid (D220), almost perfectly superimposes with the dyads of each of the other DH structures (**Figure 3-29b**). A hydrogen bonding network beginning with the phenolic hydroxyl of Y164 may increase the pK_a of the active site aspartate, thereby allowing it to act as a general acid for the dehydration reaction. A similar hydrogen bonding network is conserved in five of the six known DH structures, which employ either a glutamine or histidine (H224 in RifDH10) to bridge the network between Y164 and the catalytic aspartate. The exception is the CurH DH, in which an arginine occupies the corresponding site.

DISCUSSION

RifDH10 is only the second PKS DH domain for which both the protein structure and the stereospecificity of the dehydration reaction have been experimentally determined (Valenzano *et al.*, 2010). Consistent with the high degree of similarity in overall protein architecture between RifDH10 and EryDH4, as well as the essentially identical relative positioning of their conserved active site histidine and aspartate residues, we have established that both dehydratases catalyze syn elimination of water to give *trans*-unsaturated, trisubstituted carboxylic acid thioesters. On the other hand, these two closely related dehydratases actually catalyze the dehydration of effectively enantiomeric syn-2-methyl-3-hydroxyacyl substrates. Catalysis of each dehydration

reaction is believed to involve deprotonation of C2 of the substrate by the imidazole side chain of H50 that acts as the general base, with either stepwise or concerted elimination of water promoted by protonation of the C3 hydroxyl by the carboxylic acid side chain of D220 (**Figure 3-29b**). These conserved active site residues have been shown to be essential not only in EryDH4 (Bevitt *et al.*, 1993) and in the dehydratase domain of animal FAS (Pasta *et al.*, 2007) but also in the corresponding type II FabZ dehydratase proteins of bacterial fatty acid biosynthesis (Kimber *et al.*, 2004). The topological relationship of the active site histidine and aspartate residues is consistent only with net syn elimination of water from the substrate. Crotonase (enoyl-CoA hydratase), a mechanistically closely related enzyme from the fatty acid β-oxidation pathway, although differing from PKS and FAS DH domains in both active site residues and overall protein structure, has also been shown to catalyze reversible concerted syn addition of water to *trans*-enoyl-CoA substrates (Agnihotri & Liu, 2003; Bahnson *et al.*, 1991).

In spite of all these mechanistic and structural similarities to other characterized DH domains, RifDH10 is unique in catalyzing the diastereospecific syn dehydration of a (2*S*,3*S*)-2-methyl-3-hydroxyacyl-RifACP10 substrate. This is in contrast to the demonstrated specificity of EryDH4, TylDH2, and NanDH2, all of which utilize only (2*R*,3*R*)-2-methyl-3-hydroxyacyl-ACP substrates, as well as the specificity of the DH domain of yeast FAS for (3*R*)-hydroxyacyl-ACP substrates (Sedgwick *et al.*, 1978). Intriguingly, the intrinsic preference of RifDH10 for the (2*S*,3*S*)-2-methyl-3-hydroxyacyl thioesters is only evident when the substrate is tethered to its native acyl carrier protein, RifACP10. By contrast, RifDH10-catalyzed dehydration of the corresponding

noncognate EryACP6 as well as the NAC and pantetheine thioester analogues is specific for the corresponding (2*R*,3*R*)-2-methyl-3-hydroxyacyl diastereomers, while in no case do either the (2*R*,3*S*)- or (2*S*,3*R*)-2-methyl-3-hydroxyacyl derivatives undergo dehydration. This reversal of substrate specificity is unprecedented. Although several EryKS domains have been shown to exhibit strong preferences for their cognate ACP domains in both intramodular condensation reactions and intermodular chain transfers (Kim *et al.*, 2004; Kapur *et al.*, 2012; Guo *et al.*, 2012), and the stereoselectivity of KR-catalyzed reductions can require or be enhanced by tethering of 3-ketoacyl thioester substrates to an ACP domain (Valenzano *et al.*, 2009; Siskos *et al.*, 2005), this is the first example of the complete reversal of the diastereospecificity of a biochemical reaction based on the nature of the thioester conjugate.

Several interlocking lines of argument support the biosynthetic relevance of the observed strict preference for the RifACP10-linked, (2*S*,3*S*)-2-methyl-3-hydroxyacyl thioester substrate. (1) RifACP10 is derived from the same parent module of the rifamycin synthase as is RifDH10 and therefore must represent the native carrier for dehydration of the natural ACP-bound acyclic undecaketide substrate within Rif module 10. (2) RifKR10, the paired KR domain that is also harbored in Rif module 10, generates exclusively the (2*S*,3*S*)-2-methyl-3-hydroxyacyl thioester, independent of the nature or origin of the attached ACP domain. Under natural conditions, therefore, RifDH10 should encounter only the (2*S*,3*S*)-2-methyl-3-hydroxyacyl undecaketide that will be tethered to RifACP10. (3) While most of the substrate analogues that have been tested *in vitro* with RifDH10 are diketide derivatives, longer chain acyl thioester substrates have no effect on

the observed stereospecificity of RifDH10. (4) It highly unlikely that the *in vitro* results are artifacts and that the natural reaction on the undecaketide substrate would generate a *cis* double bond on the basis of both the highly conserved active site geometry of RifDH10, including the precise positioning of the catalytic histidine and aspartate residues in a topology that has been shown to catalyze syn dehydration by several other DH domains, and the exclusive formation and rehydration of *trans*-α,β-unsaturated trisubstituted enoyl thioesters. (5) If these *in vitro* results were simply the spurious consequence of the use of a deconstructed *in vitro* enzyme system, one would expect to see a simple degradation in overall diastereospecificity, rather than the observed complete reversal of the diastereospecificity that is cleanly correlated with the use of RifDH10 within its natural context.

The syn stereochemistry of the DH-catalyzed dehydration reaction is thus an intrinsic catalytic property that is rooted in the topological placement of the active site histidine and aspartate residues. This reaction stereochemistry should not be influenced by simple variations in the structure and stereochemistry of the substrate. While it is not unusual to observe degradation in substrate stereospecificity with the use of alternative substrates, to the best of our knowledge it is unprecedented to encounter a complete reversal in the apparent diastereospecificity that is a consequence of a change in the nature of the thioester conjugate. Although both the (2*S*,3*S*)- and (2*R*,3*R*)-2-methyl-3-hydroxyacyl thioesters can be modeled into the RifDH10 active site, with productive alignments of both the syn-related C2 proton and the C3 hydroxyl group with the corresponding H50 and D220 residues, the attached pantetheinyl moieties would be

differently positioned in each case (**Figure 3-30**). While there is as yet no protein structure of a DH−ACP complex, several lines of evidence have supported the postulated existence of strong interactions between DH domains and their cognate ACP domains. Both pull-down and surface plasmon resonance experiments have shown a high affinity of the Helicobacter pylori ACP (HpACP) for the HpFabZ protein, with a K_d of 10 nM (Liu *et al.*, 2007). Similarly, *in silico* protein ligand docking and molecular dynamics simulations have been used to model complexes involving specific interactions of EryDH4 and holo-EryACP4 (Anand & Mohanty, 2012). We surmise that substitution by the noncognate EryACP6, which has a sequence that is only 43% identical to that of RifACP10, or by the simpler NAC and pantetheine analogues results in the loss of critical RifDH10-RifACP10 interactions that are essential for the proper positioning of the tethered substrate within the active site, thereby permitting syn dehydration of the anomalously bound, diastereomeric (2*R*,3*R*)-2-methyl-3-hydroxyacyl thioester analogues.

Our results strongly indicate that during the course of rifamycin biosynthesis the combined action of the RifKR10 and RifDH10 domains from the terminal module of the rifamycin PKS naturally generates an acyclic (*E*)-2-methyl-2-enoyl undecaketide. We therefore suggest that isomerization of the (*E*)-2,3-double bond of the initially formed RifACP10-bound undecaketide intermediate to the corresponding *cis* geometry that is characteristic of both the cyclized intermediate proansamycin X and the ultimately derived rifamycin B probably most likely occurs during the subsequent macrolactamization reaction that is catalyzed by the discrete amide synthase RifF.

MATERIALS AND METHODS

Isopropyl thio-β-D-galactopyranoside (IPTG) was purchased from Invitrogen. All other chemical reagents were purchased from Sigma-Aldrich and utilized without further purification. Ni-NTA affinity resin was purchased from Qiagen. Amicon Ultra Centrifugal Filter Units [Amicon Ultra-15, 30000 molecular weight cutoff (MWCO)] were purchased from Millipore. DNA primers were synthesized by Integrated DNA Technologies. Recombinant Ery[KS6][AT6], EryACP6, RifKR7, RifKR10, TylKR1, EryKR1, EryKR6, PicTE, and Sfp were each expressed and purified as previously described (You *et al.*, 2013; Chen *et al.*, 2006; Chen *et al.*, 2007; Kim *et al.*, 2004). Reference standards for chiral GC−MS analysis of methyl (2*S*,3*S*)-2-methyl-3-hydroxypentanoate (4a), methyl (2*R*,3*R*)-2-methyl-3-hydroxypentanoate (4b), methyl (2*R*,3*S*)-2-methyl-3-hydroxypentanoate (4c), and methyl (2*S*,3*R*)-2-methyl-3-hydroxypentanoate (4d) were each prepared synthetically or chemoenzymatically as previously described (You *et al.*, 2013; Valenzano *et al.*, 2009). (*E*)-2-Methyl-2-pentenoic acid (3) was purchased from Sigma-Aldrich and used to prepare (*Z*)-2-methyl-2-pentenoic acid as previously described (Valenzano *et al.*, 2010). Propionyl-N-acetylcysteamine thioester and (2*S*,3*R*)-2-methyl-3-hydroxypentanoyl-SNAC were prepared as described previously (Valenzano *et al.*, 2009). The synthesis of acyl-*S*-NAC and acyl-*S*-pantetheine thioester substrates and products is described in the Supporting Information. (2*RS*)-2-Methyl-3-ketopentanoyl-CoA was prepared by L. Collett Pilcher as previously described (You *et al.*, 2013; Wu *et al.*, 2001; Haapalainen *et al.*, 2007; Agnihotri & Liu, 2003), purified by reversed-phase C-18 high-performance liquid

chromatography (HPLC), and characterized by ^{1}H nuclear magnetic resonance and electrospray ionization mass spectrometry (ESI-MS). Synthetic genes encoding recombinant domains RifDH10 and RifACP10, optimized for expression in *Escherichia coli*, were designed and prepared by DNA2.0 and supplied in vector pJ201.

General methods were as previously described (Valenzano *et al.*, 2009; Sambrook *et al.*, 1989). Growth media and conditions used for *E. coli* and standard methods for handling *E. coli in vivo* and *in vitro* were those described previously, unless otherwise noted (Sambrook *et al.*, 1989). All DNA manipulations were performed following standard procedures (Sambrook *et al.*, 1989). DNA sequencing was conducted at the University of California Davis Sequencing Facility (Davis, CA). All proteins were handled at 4 °C unless otherwise stated. Protein concentrations were determined according to the method of Bradford (Bradford, 1976), using Hewlett-Packard 8452A Diode Array or Thermo Evolution Array UV−vis spectrophotometers with bovine serum albumin as the standard. Sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) and DNA gels were imaged and analyzed with a Bio-Rad ChemiDoc MP System. A Bio-Rad FX-Pro Plus Molecular Imager was utilized for radio thin layer chromatography (TLC) analysis. GC−MS analysis was performed on a GC−MS Hewlett-Packard Series 2 GC-MSD instrument (70 eV EI) in positive ion mode using a capillary CP-Chirasil-Dex CB column (25 m × 0.32 mm) from Agilent Technologies. For resolution and analysis of (*E*)-2-methyl-2-pentenoic acid (3) and (*Z*)-2-methyl-2-pentenoic acid, three different programs were used: method 1, initial oven temperature of 50 °C for 1 min, increased to 200 °C at a rate of 7.5 °C/min; method 2, initial oven

temperature of 50 °C for 1 min, increased to 200 °C at a rate of 15 °C/min; method 3, initial oven temperature of 65 °C for 1 min, increased to 100 °C at a rate of 0.5 °C/min and then to a final temperature of 200 °C at a rate of 20 °C/min. For resolution and analysis of the four diastereomers of methyl 2-methyl-3-hydroxypentanoate (4), the following program was used: method 4, initial oven temperature of 50 °C for 1 min, increased to 90 °C at a rate of 1 °C/min and then to 200 °C at a rate of 20 °C/min. A Thermo LXQ instrument equipped with a Surveyor HPLC system and Waters Symmetry C18 column (2.1 mm × 50 mm, 3.5 μm) was utilized for HPLC−ESI-MS analysis in positive ion mode.

General cloning strategy

The design of rifamycin module 10 domain boundaries for cloning of RifDH10, RifDH10-RifKR10, and RifACP10-NusA is described in the Supporting Information (**Figures 3-3**, **3-4**, and **3-6**). Synthetic genes encoding RifDH10, RifDH10-RifKR10, and RifACP10-NusA (De Marco, 2004), optimized for expression in *E. coli* and flanked by suitable restriction sites, were ligated into pET28a, and the resultant plasmids were used to transform the expression host *E. coli* BL21(DE3).

RifDH10 expression and purification

The synthetic gene for RifDH10, flanked by 5′-NdeI and 3′-XhoI sites, was ligated into pET28a. *E. coli* BL21(DE3) transformed with the RifDH10 expression plasmid was inoculated into Luria-Bertani medium containing 50 μg/L kanamycin at 37 °C, grown to an OD_{600} of 0.4, and induced with 0.5 mM IPTG. After 12 h at 15 °C, cells

were collected by centrifugation and resuspended in lysis buffer [0.5 M NaCl, 10% (v/v) glycerol, and 0.1 M 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) (pH 7.5)]. Following sonication, cell debris was removed by centrifugation (30000g for 30 min). The supernatant was poured over a column of nickel-NTA resin (Thermo Scientific), which was then washed with 50 mL of lysis buffer containing 15 mM imidazole and eluted with 5 mL of lysis buffer containing 150 mM imidazole. RifDH10 was further purified using a Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated with 150 mM NaCl and 10 mM HEPES (pH 7.5). The eluted protein (~100 mg of purified RifDH10 from 2 L of culture) was concentrated to 15 mg/mL in the equilibration buffer and stored at −80 °C until the protein was needed. SDS−PAGE analysis of purified His_6RifDH10 protein showed an Mr of 34900 (calcd molecular mass of 34109 Da) (**Figure 3-5**).

Apo-RifACP10-NusA fusion protein

The synthetic gene for RifACP10, flanked by 5′-NdeI and 3′-XhoI sites, was initially ligated into the NdeI and XhoI sites of pET28a. Because the resultant recombinant RifACP10 was obtained only as insoluble inclusion bodies when expressed in *E. coli* BL21(DE3), we constructed the corresponding RifACP10-NusA protein. The synthetic DNA encoding RifACP10 was amplified via polymerase chain reaction from the RifACP10 construct using the primer pair of pET28a-FP (5′-ATGGGCAGCAGCCATCATCAT-3′) and RifACP10-EcoRI-RP [5′-TATTTAAAATTCAGAATTCCAGCAGCTCATTCAAGTGTGC-3′ (EcoRI site in

italics)]. The resultant amplified DNA was digested with NdeI and EcoRI and ligated into the previously described plasmid (You *et al.*, 2013), RifKR10-NusA-pET28, that had been digested with NdeI and EcoRI to excise the DNA segment encoding RifKR10 and to replace it with DNA encoding RifACP10, thus giving apo-RifACP10-NusA-pET28 that was used to transform *E. coli* BL21(DE3). Recombinant His_6 tag-apo-RifACP10-NusA, harboring an HRV 3C protease cleavage site between RifACP10 and the C-terminal NusA, was expressed in soluble form by growth in Terrific Broth or in Super Broth until the OD600 reached 0.4−0.8. Protein expression was induced by addition of 0.2 mM IPTG, and the cell culture was continuously grown for an additional 16− 18 h at 18 °C overnight. The cells were harvested by centrifugation at 4200g for 20 min, and the cell pellet was collected and stored at −80 °C. The frozen cells were thawed at room temperature (RT) and dissolved and lysed on ice for 30 min in starting buffer [500 mM NaCl and 50 mM phosphate (pH 7.8)] containing 1 mg/mL lysozyme, followed by sonication. The cell supernatant and the pellet were separated by centrifugation at 23000g for 30 min, and the supernatant was loaded on a precharged Ni-NTA column. The column was washed with starting buffer and then washing buffer [300 mM NaCl, 50 mM phosphate (pH 7.6), and 10 mM imidazole], and proteins were eluted from the Ni column with elution buffer [150 mM NaCl, 50 mM phosphate (pH 7.5), 150 mM imidazole, and 10% glycerol]. The eluted fractions were collected and concentrated with an Amicon MWCO 30000 filter, and buffer was exchanged with a PD-10 column with 150 mM NaCl, 50 mM phosphate (pH 7.5), and 10% glycerol, yielding ∼200 mg of His_6 tag-apo-RifACP10-NusA/2 L of culture. SDS−PAGE analysis of purified His_6 tag-apo-

RifACP10-NusA protein showed an Mr of 70900 (calcd molecular mass of 68955 Da) (**Figure 3-5**).

RifDH10-KR10 fusion protein

DNA encoding the synthetic gene for RifKR10 was amplified from the previously described expression construct (You *et al.*, 2013) using the primer pair of RifKR10-NdeI-BamHI-FP [5′-ATTTTCGAATTACATATGATGAATTTTAAAGGATCCCTGTACCGCGTCGACTG-3′ (NdeI site in bold and BamHI site in italics)] and pET-28a XhoI RP (5′-TCGGGCTTTGTTAGCAGC-3′). The amplified DNA fragment was digested with NdeI and XhoI followed by ligation into the pET-28a vector to give RifKR10-NdeI-BamHI-pET28a. DNA encoding RifDH10 was amplified from the expression construct described above using the primer pair of RifDH10-PAGA-BamHI-RP [5′-TAATTCGAAAATGGATCCTGCACCCGCTGGACCCGCTGCAGTGGTC-3′ (BamHI site in italics)] and pET28a-FP (5′-ATGGGCAGCAGCCATCATCAT-3′) followed by DNA restriction by NdeI and BamHI. The resultant DNA fragment was ligated into NdeI and BamHI-digested plasmid RifKR10-NdeI-BamHI-pET28a. The His_6 tag-RifDH10-KR10 protein was expressed in *E. coli* BL21(DE3) and purified by the same procedures used to express the His_6 tag-apo-RifACP10-NusA protein and then further purified by size exclusion chromatography, yielding 16 mg of partially purified His_6 tag-Rif[DH10][KR10] didomain protein/2 L culture. SDS−PAGE analysis of partially

purified His_6 tag-Rif[DH10][KR10] protein gave an Mr of 80100 (calcd molecular mass of 80150 Da) (**Figure 3-5**).

Incubation of RifDH10 with (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-RifACP10-NusA

A mixture of 500 μM (2*R*S)-2-methyl-3-ketopentanoyl-CoA, 40 μM Sfp (surfactin phosphopantetheinyl transferase) (Weinreb *et al.*, 1998), and 400 μM apo-RifACP10-NusA was added to reaction buffer [150 mM NaCl and 50 mM NaH_2PO_4 (pH 7.2)] containing 20 mM $MgCl_2$ and 4 mM DTT. After incubation at 37 °C for 1 h, 300 μM RifKR7, 300 μM RifDH10, and 2 mM NADPH were added, and the mixture was incubated for an additional 1 h at RT. The final reaction volume was 500 μL. The RifACP10-NusA protein was hydrolytically released by incubation with 100 μM PicTE for 10 min at RT. The reaction mixtures were acidified with 1 M HCl to pH <3 and extracted with ethyl acetate. After solvent evaporation, the product was dissolved in 100 μL of methanol and the unsaturated diketide was analyzed by chiral GC−MS and direct comparison with authentic standards of (*E*)-2-methyl-2-pentenoic acid and (*Z*)-2-methyl-2-pentenoic acid, using both methods 1 and 2 (**Figure 3-8**) (Valenzano *et al.*, 2010). In control experiments conducted in the absence of RifDH10, 20 μL of TMS-diazomethane was added to the hydrolyzed organic extract and the derived diketide methyl ester was analyzed by chiral GC−MS and direct comparison with authentic methyl (2*S*,3*S*)-2-methyl-3-hydroxypentanoate, using method 4 (**Figure 3-9**). As previously reported, direct comparison with authentic standards 4a and 4b was essential because the observed retention time on the chiral column varied slightly from run to run, although the order of

elution never changed (You *et al.*, 2013; Valenzano *et al.*, 2009). Incubations with TylKR1, EryKR6, and EryKR1 in place of RifKR7 were also conducted to generate the corresponding RifACP10-bound intermediates (2*R*,3*R*)-1b, (2*R*,3*S*)-1c, and (2*S*,3*R*)-1d. The derived hydrolysis products were analyzed by GC−MS and direct comparison with authentic standards of each diastereomer of methyl 2-methyl-3-hydroxypentanoate (4a−d) (method 4) and both (*E*)-2-methyl-2-pentenoic acid (3) and (*Z*)-2-methyl-2-pentenoic acid (method 2), as described for the incubations using RifKR7 (**Figures 3-10**, **3-11**, and **3-12**).

Incubation of the recombinant RifDH10-RifKR10 didomain with chemoenzymatically generated (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10

A mixture of 500 μM (2*R*S)-2-methyl-3-ketopentanoyl-CoA, 40 μM Sfp, and 400 μM apo-RifACP10-NusA was added to reaction buffer [150 mM NaCl and 50 mM NaH_2PO_4 (pH 7.2)] containing 20 mM $MgCl_2$ and 4 mM DTT. After incubation at 37 °C for 1 h, 300 μM RifDH10-RifKR10 and 2 mM NADPH were added, and the mixture was incubated for an additional 1 h at RT. The final reaction volume of each assay was 500 μL. The diketide acid was hydrolytically released from RifACP10-NusA by incubation with 100 μM PicTE for 10 min at RT. The reaction mixtures were acidified with 1 M HCl to pH <3 and extracted with ethyl acetate. After solvent evaporation, the product was dissolved in 100 μL of methanol and the formation of unsaturated (*E*)-2-methyl-2-pentenoic acid was confirmed by GC−MS (method 2) and direct comparison with authentic standards of 3 and (*Z*)-2-methyl-2-pentenoic acid. For the detection of the

reduced diketide product, 20 μL of TMS-diazomethane was added and the derived diketide methyl ester, methyl (2*S*,3*S*)-2-methyl-3-hydroxypentanoate (4a), was identified by chiral GC−MS (method 4) and direct comparison with authentic standards (**Figure 3-13**).

Incubation of recombinant RifDH10 with chemoenzymatically generated (*E*)-2-methyl-2-pentenoyl-Rif-ACP10 and (*E*)-2-methylpentenoyl-EryACP6.

A mixture of 500 μM (*E*)-2-methyl-2-pentenoyl-CoA, 40 μM Sfp, and either 400 μM apo-RifACP10-NusA or 200 μM apo-EryACP6 was added to reaction buffer [150 mM NaCl and 50 mM NaH_2PO_4 (pH 7.2)] containing 20 mM $MgCl_2$ and 4 mM DTT. After incubation at 37 °C for 1 h, 300 μM RifDH10 was added and the mixture was incubated for an additional 1 h at RT. The final reaction volume of each assay was 500 μL. The RifACP10-NusA or EryACP6 was hydrolytically released by incubation with 100 μM PicTE for 10 min at RT. The reaction mixtures were acidified with 1 M HCl to pH <3 and extracted with ethyl acetate. After solvent evaporation, the product was dissolved in 100 μL of methanol and 20 μL of TMS-diazomethane was added. The derived diketide methyl esters were analyzed as described above by chiral GC−MS and direct comparison with authentic standards [**Figure 3-14** (RifACP10) and **Figure 3-19** (EryACP6)]. In a control incubation without the added RifDH10 domain, the chemoenzymatically generated 2-methyl-2-pentenoyl-RifACP10-NusA was treated with HRV 3C protease to remove the C-terminal NusA and the molecular mass of the resulting 2-methyl-2-pentenoyl-RifACP10 was verified by LC−ESI(+)-MS, which gave an

observed value of 13144 Da (calcd for [M + H]+ ion of 2-methyl-2-pentenoyl-RifACP10, 13143.9 Da) (**Figure 3-31**). LC−ESI(+)-MS was also used to confirm the formation of 2-methyl-2-pentenoyl-ACP6 (6), with a molecular mass (observed) of 11728 Da (calcd 11727.7 Da), compared to the peak for holo-eryACP6, with a molecular mass (observed) of 11632 Da (calcd 11631.6 Da), with the difference of 96 Da matching that expected for addition of the 2-methyl-2-pentenoyl moiety. Additional control incubations established that the isomeric (*Z*)-2-methylpentenoyl-EryACP6 was not a substrate for RifDH10.

TLC−phosphorimaging assay of the incubation of recombinant RifDH10 with reconstituted Ery[KS6]-[AT6], EryACP6, and recombinant KR domains

A radio-TLC assay of the conversion of propionyl-SNAC to 2-methyl-3-hydroxypentanoic acids and 2-methyl-2-pentenoic acid by coupled incubation with reconstituted Ery[KS6][AT6] and EryACP6 in combination with EryKR6, TylKR1, EryKR1, or RifKR7 and RifDH10 was conducted as previously described (Valenzano *et al.*, 2010; You *et al.*, 2013; Valenzano *et al.*, 2009). For each incubation, 5 mM propionyl-SNAC and 40 μM Ery[KS6][AT6] were added to reaction buffer [50 mM NaH_2PO_4 (pH 7.2)] containing 2.5 mM TCEP and the mixture was incubated at RT. After 1 h, 200 μM holo-EryACP6, 300 μM methylmalonyl-CoA containing a trace amount of [2-14C]methylmalonyl-CoA, 2 mM NADPH, and individual KR domains (300 μM) with or without 300 μM RifDH10 were added to the reaction mixture, which was then incubated for an additional 1 h. The final reaction volume of each assay was 125 μL. The reaction was quenched, and the diketide acids were hydrolytically released from

EryACP6 by addition of 0.5 M NaOH (40% by volume) followed by incubation at 65 °C for 20 min. After adjustment to pH <3 with 1 M HCl and extraction with ethyl acetate, the concentrated organic extracts were dissolved in 15 μL of ethyl acetate and spotted on a silica gel TLC plate, which was developed with a 1:9 dichloromethane/ ethyl acetate mixture containing 0.1% AcOH. The dried TLC plate was then visualized by phosphorimaging for 48 h (**Figure 3-16**).

Incubation of recombinant RifDH10 or the RifDH10-RifKR10 didomain with reconstituted Ery[KS6][AT6], EryACP6, and TylKR1

The incubation procedures for the *in situ* generation of (2*R*,3*R*)-2-methyl-3-hydroxyacyl-EryACP6 derivatives were based on those previously described (You *et al.*, 2013; Valenzano *et al.*, 2009). In a typical assay, 5 mM propionyl-SNAC or 5 mM (2*S*,3*R*)-2-methyl-3-hydroxypentanoyl-SNAC was preincubated with 40 μM Ery[KS6][AT6] in 50 mM NaH_2PO_4 (pH 7.2) containing 2.5 mM tris-2-carboxyethylphosphine (TCEP) at RT. After 1 h, 200 μM holo-EryACP6, 300 μM methylmalonyl-CoA, 2 mM NADPH, 300 μM recombinant TylKR1, and 300 μM recombinant RifDH10 were added to the reaction mixture, which was then incubated for an additional 1 h (**Figure 3-17**). In a second experiment, 300 μM Rif[DH10][KR10] didomain was added instead of the TylKR1 and RifDH10 pair (**Figure 3-28**). The final reaction volume of each assay was 500 μL. The reaction was quenched, and EryACP6 was hydrolytically released from the product by addition of 0.5 M NaOH (40% by volume) and incubation at 65 °C for 20 min. Alternatively, EryACP6 was enzymatically

released by incubation with 100 μM PicTE for 10 min at RT. The reaction mixtures were acidified with 1 M HCl to pH <3 and then extracted with ethyl acetate. After solvent evaporation, the product was dissolved in 100 μL of ethyl acetate or methanol and analyzed by chiral GC−MS (method 3) (**Figures 3-17** and **3-28**). TMS-diazomethane (20 μL) was added to convert recovered 2-methyl-3-hydroxypentanoates to the derived 2-methyl-3-hydroxypentanoate methyl esters that were analyzed by chiral GC−MS and direct comparison with synthetic standards (method 4) (Valenzano *et al.*, 2009). For the GC−MS analysis of methyl (2*E*,4R,5R)-2,4-dimethyl-5-hydroxy-2-heptenoate, the temperature program of method 3 was used (**Figure 3-18**).

Incubation of RifDH10 with acyl-*S*-NAC or pantetheine thioesters

RifDH10 (30 μM) was incubated overnight with each of the chemoenzymatically prepared 2-methyl-3-hydroxyacyl or 2,3-unsaturated enoyl-*S*-NAC or pantetheine thioester analogues (at 10 mM) in a 25 μL solution containing 150 mM NaCl, 10% (v/v) glycerol, and 150 mM HEPES (pH 7.5). After 16 h, the reaction mixture was injected onto a C18 reversed-phase HPLC column and monitored over a range of λ values of 200−600 nm (100% water with 0.1% TFA to 100% MeOH with 0.1% TFA, over 30 min). All products were characterized by mass spectrometry and comparison with authentic samples as appropriate (**Figures 3-20**, **3-21**, **3-22**, **3-23**, **3-24**, **3-25**, **3-26**, and **3-27**).

RifDH10-catalyzed dehydration of (2*R*,3*R*)-2-methyl-3-hydroxybutanoyl-*S*-pantetheine and subsequent acyl transthioesterification to NAC.

trans-2-Methyl-2-butenoyl-*S*-pantetheine (11) (20 mg) was incubated overnight in 150 mM NaCl, 150 mM HEPES (pH 7.5), 10% (v/v) glycerol, and 1 mg/mL RifDH10 in a total volume of 1 mL at 22 °C. After 16 h, the reaction mixture was injected onto a C18 reversed-phase HPLC column and monitored over a range of λ values of 200− 600 nm (100% water with 0.1% TFA to 100% MeOH with 0.1% TFA, over 15 min). A peak that was not observed in a control reaction without RifDH10 with a λ_{max} of 232 nm was collected and concentrated. The concentrated fractions yielded 5 mg of the hydrated product, (2*R*,3*R*)-2-methyl-3-hydroxybutanoyl-*S*-pantetheine (12b). To confirm its stereochemistry, 3 mg of (2*R*,3*R*)-2-methyl-3-hydroxybutanoyl-*S*-pantetheine (12b) was resuspended in 5 mL of 50 mM NaHCO3 (aqueous) and 50 mM NAC and stirred overnight. The reaction mixture was injected onto a C18 reversed-phase HPLC column and monitored at a range of λ values of 200−600 nm (100% water with 0.1% TFA to 100% MeOH with 0.1% TFA, over 15 min). A peak that was not observed in a control reaction without NAC with a λ_{max} of 232 nm was collected and concentrated. The compound ran in a manner identical to that of authentic (2*R*,3*R*)-2-methyl-3-hydroxybutanoyl-*S*-NAC (9b) on a ChiralCel OC-H column (250 mm × 4.6 mm) on a Beckman Coulter HPLC system equipped with a 20 μL loop with an isocratic flow rate of 0.6 mL/min using a 1:24 mixture of ethanol and hexanes as the mobile phase (**Figure 3-27**) (Piasecki *et al.*, 2011). Compounds were observed by 235 nm UV absorbance.

Synthetic standards, prepared as previously described, were provided by S. Piasecki (Piasecki *et al.*, 2011).

Crystallization and structure determination

Crystals of RifDH10 grew over a period of 2 days to 1 week by sitting drop vapor diffusion at 22 °C. Drops were formed by mixing 3 μL of a protein solution [15 mg/mL RifDH10, 150 mM NaCl, and 10 mM HEPES (pH 7.5)] with 1 μL of crystallization buffer [1.1 M sodium citrate and 0.1 M HEPES (pH 6.8)]. Crystals were soaked briefly in crystallization buffer with 15% (v/v) glycerol before being frozen in liquid nitrogen. The diffraction data, collected at ALS beamline 5.0.2, were processed with iMosflm and scaled with SCALA from the CCP4 suite (Collaborative Computational Project Number 4, 1994). The structure was determined to 1.82 Å resolution by molecular replacement with PhaserMR (McCoy *et al.*, 2007) using EryDH4 [Protein Data Bank (PDB) entry 3EL6] as the search model (Keatinge-Clay, 2008). The model generated from the molecular replacement solution was refined with Coot (Emsley *et al.*, 2010) and Refmac5 (Murshudov *et al.*, 1997). Aside from the N- and C-termini, only loop residues 232−242 were not shown in the electron density maps.

NMR and LC-MS characterization of *S*-pantetheine esters

^{1}H NMR analysis of *S*-pantetheine esters was performed on a Varian Mercury 400 MHz instrument. LC-MS was performed on an Agilent Technologies 1200 Series HPLC with a Gemini C_{18} column (5 μm, 2 × 50 mm; Phenomenex) coupled to an Agilent Technologies 6130 quadrupole mass spectrometer system equipped with an electrospray-

ionization source. A 5–95% B gradient over 12 min at a flow rate of 0.7 mL/min was run in which the mobile phases were (A) H_2O with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid.

N-Acetylcysteamine (NAC), Meldrum's acid derivatives, the methylketene dimer, (2*RS*)-2-methyl-3-ketobutanoyl-*S*-NAC, (2*RS*)-2-methyl-3-ketopentanoyl-*S*-NAC, and (3*R*)-3-hydroxypentanoyl-*S*-NAC were synthesized as previously reported (Piasecki *et al.*, 2011), as was *trans*-2-butenoyl-*S*-pantetheine (Zheng *et al.*, 2012).

(2*RS*)-2-methyl-3-ketopentanoyl-*S*-pantetheine

Sodium borohydride (80 mg, 16 eq.) was added to 94 mg D-pantethine (1 eq.) dissolved in 5 mL 80:20 MeOH:0.25 M $NaHCO_3$ (aq.) over 20 min at 22 °C. After 1 h, the reaction was quenched by adding glacial acetic acid dropwise until bubbling ceased and then was buffered with 250 mM HEPES (pH 7.5). 11.2 mg (1.4 eq.) of methylketene dimer was added to the solution at 22 °C. After 1 h, the reaction was evaporated and salts were removed using a plug of silica gel (2% MeOH:dichloromethane).

(2RS)-2-Methyl-3-ketopentanoyl-S-pantetheine: ^{1}H NMR (400 MHz, D_2O): 0.71 (s, 3H), 0.74 (s, 3H), 0.81-0.88 (t, 3H, J = 8 Hz), 1.20-1.25 (d, 3H, J = 8 Hz), 2.28-2.33 (m, 2H) 2.51-2.58 (q, 2H, J = 8 Hz), 2.90-2.97 (t, 2H, J = 7.2 Hz), 3.20-3.26 (m, 2H), 3.30-3.40 (m, 2H), 3.40-3.45 (m, 2H), 3.81 (s, 1H), 4.00-4.05 (q, 1H, J = 8 Hz). ESI-MS expected mass: 391.5, observed mass: 391.2.

(2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-*S*-pantetheine, (2*R*,3*S*)-2-methyl-3-hydroxypentanoyl-*S*-pantetheine, (2*S*,3*R*)-2-methyl-3-hydroxypentanoyl-*S*-pantetheine, and (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-*S*-pantetheine

Each of the stereoisomers of 2-methyl-3-hydroxypentanoyl-*S*-pantetheine was prepared from the incubation of (2*RS*)-2-methyl-3-ketopentanoyl-*S*-pantetheine with various recombinant KR domains that have been shown to stereospecifically reduce the corresponding NAC-bound compounds (TylKR1, AmpKR2, EryKR1, and AmpKR1, respectively).[1] Reduction reactions were carried out in 10% (v/v) glycerol, 150 mM HEPES (pH 7.5), 100 mM NaCl, 200 mM D-glucose, 500 μM $NADP^+$, 10 μM KR, 1 μM GDH (cloned from *Bacillus subtilis*), and 2-10 mM (2*RS*)-2-methyl-3-ketopentanoyl-*S*-pantetheine in a total volume of 200 μL for 24 h at 22 °C. The reaction was monitored by TLC, and the reduction product was purified by silica gel chromatography with 1:1 ethyl acetate:acetone.

***trans*-2-methyl-2-butenoyl-*S*-pantetheine (11)**

D-pantethine (1 mL of 0.5 M aq. soln) was reduced to D-pantetheine by addition of 75 mg DTT (1 eq.) and stirring for 1 h at 22 °C. Water was removed by vacuum to leave a clear, viscous liquid that was resuspended in 10 mL DCM. *trans*-2-Methyl-2-butenoic acid (100 mg, 2.1 eq.) (TCI Chemicals), 150 mg EDC (2.0 eq.) (Alfa Aesar), and 15 mg DMAP (0.25 eq.) (Acros Organics) were dissolved in 2 mL DCM and added to the pantetheine solution. The reaction was stirred at 0 °C for 1 h and 22 °C for an additional hour. All solvent was removed by vacuum, leaving a clear, viscous liquid that

was purified by silica gel chromatography with 1:1 ethyl acetate:acetone. *trans*-2-*Methyl-2-butenoyl-S-pantetheine* ***(11)***: ^{1}H NMR (400 MHz, D_2O) □ 0.71 (s, 3H), 0.74 (s, 3H), 1.66-1.70 (m, 6H), 2.29 (t, 3H, *J* = 6.5 Hz), 2.91 (t, 3H, *J* = 6.5 Hz), 3.20-3.36 (m, 8H), 3.81 (s, 1H), 6.80 (qq, 1H, *J*1 = 1.0 Hz, *J*2 = 6.6 Hz).

3*R*-hydroxybutanoyl-*S*-pantetheine

RifDH10 (30 μM) was incubated with 10 mM *trans*-2-butenoyl-*S*-pantetheine in a 1 mL solution containing 150 mM NaCl, 10% (v/v) glycerol, 150 mM HEPES pH 7.5. After 24 h, the reaction was injected on a C_{18} reversed-phase HPLC column (100% water with 0.1% TFA to 100% MeOH with 0.1% TFA, 30 min), and the peak corresponding to the hydrated species was collected and concentrated under vacuum.

3R-Hydroxybutanoyl-S-pantetheine: ^{1}H NMR (400 MHz, D_2O) □ 0.71 (s, 3H), 0.74 (s, 3H), 1.05 (d, 3H, *J* = 6.4 Hz), 2.29 (t, 3H, *J* = 6.5 Hz), 2.63 (d, 2H, *J* = 6.4 Hz), 2.89 (td, 2H, *J*1 = 2.2 Hz, *J*2 = 6.3 Hz), 3.21 (m, 3H), 3.32 (m, 3H) 3.46 (d, 1H, *J* = 4.4 Hz), 3.48 (d, 1H, *J* = 4.4 Hz), 3.81 (s, 1H), 4.09 (h, 1H, *J* = 6.5 Hz).

(2*R*,3*R*)-2-methyl-3-hydroxybutanoyl-*S*-pantetheine (12b)

Prepared in the same manner as 3*R*-hydroxybutanoyl-*S*-pantetheine, substituting *trans*-2-methyl-2-butenoyl-*S*-pantetheine for *trans*-2-butenoyl-*S*-pantetheine.

*(2R,3R)-2-Methyl-3-hydroxybutanoyl-S-pantetheine (**12b**):* ^{1}H NMR (400 MHz, D_2O) □ 0.71 (s, 3H), 0.74 (s, 3H), 0.96 (d, 3H, *J* = 7.1 Hz), 1.04 (d, 3H, 6.3 Hz), 2.29 (t, 3H, *J* = 6.5 Hz), 2.63 (dd, 3H, *J*1 = 1.2 Hz, *J*2 = 8.1 Hz), 3.20-3.36 (m, 8H), 3.81 (s, 1H).

***cis*-2-methyl-2-butenoyl-*S*-pantetheine (13)**

Prepared similarly to *trans*-2-butenoyl-*S*-pantetheine, substituting 100 mg of *cis*-2-methyl-2-butenoic acid (TCI Chemicals) for *trans*-2-methyl-2-butenoic acid. *cis*-2-*Methyl-2-butenoyl-S-pantetheine (**1**3)*: ^{1}H NMR (400 MHz, $CDCl_3$) □ 0.95 (s, 3H), 1.06 (s, 3H), 1.28 (t, 3H, *J* = 7.1 Hz), 1.96 (dq, 3H, *J*1 = 1.7 Hz, *J*2 = 7.3 Hz), 2.02 (t, 3H, *J* = 2.0 Hz), 2.44 (t, 2H, *J* = 6.3 Hz), 3.11 (q, 2H, *J* = 6.7 Hz), 3.49-3.65 (m, 4H), 4.01 (d, 1H, *J* = 4.8 Hz), 4.14 (q, 3H, *J* = 7.4 Hz), 5.93 (qq, 1H, *J*1 = 2.4 Hz, *J*2 = 7.8 Hz).

***trans*,*trans*-2,4-hexadienoyl-*S*-pantetheine**

trans,*trans*-2,4-Hexadienoic acid (1.0 g, Alfa Aesar) was stirred in 5 mL oxalyl chloride at 22 °C under a flow of nitrogen gas. After the oxalyl chloride had evaporated, reduced D-pantetheine (see synthesis of *trans*-2-methyl-2-butenoyl-*S*-pantetheine for D-pantethine reduction protocol) dissolved in 8.5 mL DCM and 1.5 mL TEA was added to the activated acid, and set to stir at 0 °C for 1 h, followed by quenching with 2 mL methanol. The reaction was washed with brine and all solvent was removed under vacuum, leaving a viscous liquid that was purified by silica gel chromatography with 1:1 ethyl acetate:acetone.

trans,trans-2,4-Hexadienoyl-S-pantetheine: ^{1}H NMR (400 MHz, D_2O): 0.71 (s, 3H), 0.74 (s, 3H), 1.12 (t, 1H, *J* = 7.3 Hz), 1.71 (d, 2H, 6.6), 2.29 (t, 3H, *J* = 6.5 Hz), 2.96 (t, 3H, *J* = 6.5), 3.20-3.36 (m, 8H), 3.81 (s, 1H), 6.16 (m, 3H), 7.14 (dd, 1H, *J* = 10.6 Hz).

ACKNOWLEDGEMENTS

Instrumentation and technical assistance at The University of Texas at Austin were provided by Dr. Art Monzingo and the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology, The University of Texas at Austin. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract DE-AC02-05CH11231. Joshua F. Detelich helped synthesize several compounds used in this chapter.

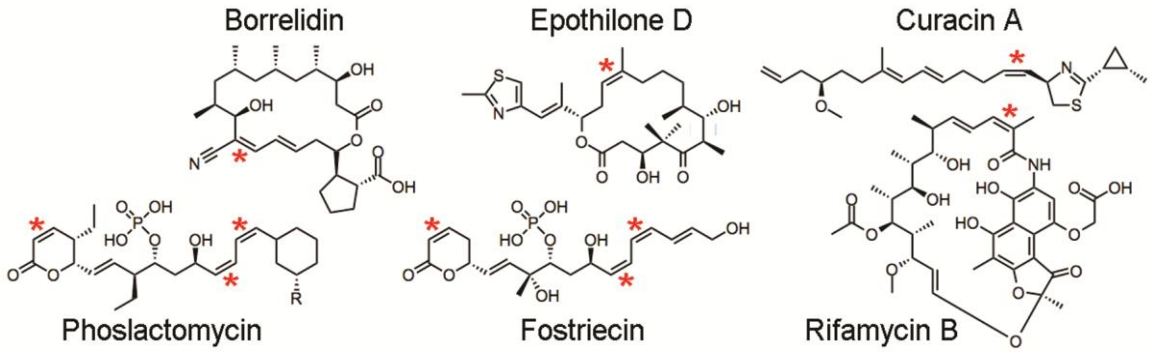


Figure 3-1. Polyketides with *cis* double bonds.

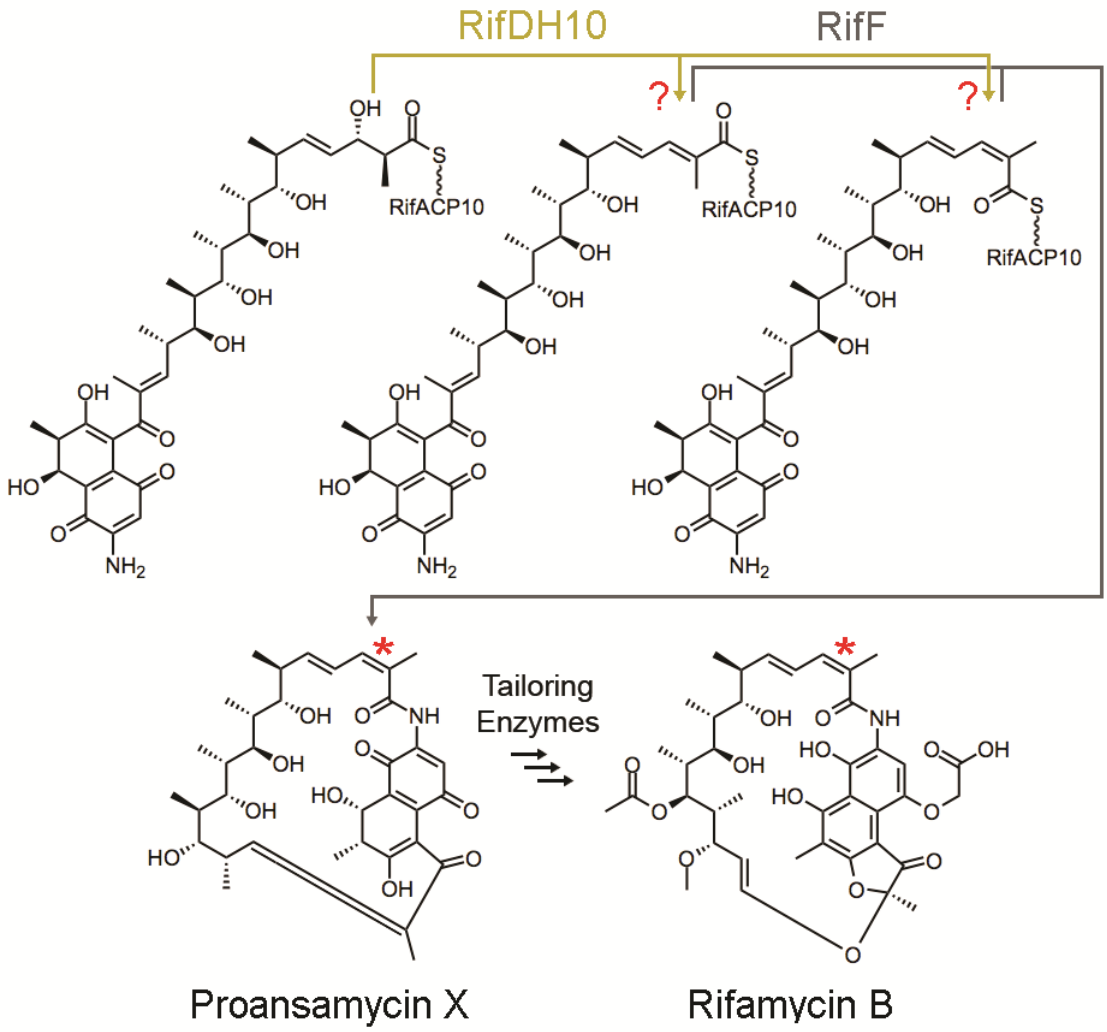


Figure 3-2. RifDH10 dehydrates a RifACP10-bound (2*S*,3*S*)-2-methyl-3-hydroxyacyl undecaketide generated by the paired ketoreductase domain RifKR10 in the terminal module 10 of the rifamycin PKS. The amide synthase RifF catalyzes the macrolactamization of the acyclic undecaketide product to the rifamycin B precursor, proansamycin X, either before or after isomerization to form the characteristic *cis* double bond.

```
   1 matdekllky lkrvtaelhs lrkqgarhad eplavvgmac rfpggvsspe dlwqlvaggv
       (.....Rifamycin PKS module 9.....)
1621 vlghagpeav radtafkdtg fdsltsvelr nrlreasglk lpatlvfdyp tpvalarylr
1681 delgdtvatt pvataaaada gepiaivgma crlpggvtdp eglwrlvrdg leglspfped
1741 rgwdlenlfd ddpdrsgtty tsrggfldga glfdagffgi sprealamdp qqrllleaaw
1801 ealegtgvdp gslkgadvgv fagvsnqgyg mgadpaelag yastagassv vsgrvsyvfg
1861 fegpavtidt acssslvamh lagqalrqge csmalaggvt vmgtpgtfve fakqrglagd
1921 grckayaega dgtgwaegvg vvvlerlsva rerghrvlav lrgsavnsdg asngltapng
1981 psqqrvirra lagaglepsd vdiveghgtg talgdpieaq allatygkdr dpetplwlgs
2041 vksnfghtqs aagvagvikm vqalrhgvmp ptlhvdrpts qvdwsagave vlteawpr
2101 ngrprragvs sfgisgtnah liieeapaep qlagpppdgg vvplvvsars pgalagqarr
2161 latflgdgpl sdvagaltsr alfgeravvv adsaeearag lgalargeda pglvrgrvpa
2221 sglpgklvwv fpgqgtqwvg mgrelleesp vfaeriaeca aalepwigws lfdvlrgdgd
2281 ldrvdvlqpa cfavmvglaa vwssagvvpd avlghsqgei aaacvsgals ledaakvval
2341 rsqaiaakls grggmasval geadvvsrla dgvevaavng pasvviagda qaldetleal
2401 sgagirarrv avdyashtrh vediedtlae alagidarap lvpflstltg ewirdegvvd
2461 ggywyrnlrg rvrfgpavea llaqghgvfv elsahpvlvq piteltdeta avvtgslrrd
2521 dgglrrllts maelfvrgve vdwtslvppa radlptyafd hehywlraad tasdavslgl
2581 agadhpllga vvqlpqsdgl vftsrlslrs hpwladhavr dvvivpgtgl velavragde
2641 agcpvldelv ieaplvvprr ggvrvqvalg gpaddgsrtv dvfslredad swlrhatgvl
2701 vpenrprgta afdfaawppp eakpvdltga ydvladvgyg ygptfravra vwrrgsgntt
2761 etfaeialpe daraeagrfg ihpalldaal hstmvsaaad tesygdevrl pfawnglrlh
2821 aagasvlrvr vakperdsls leavdesggl vvtldslvgr pvsndqltta agpagagsly
2881 rvdwtplssv dtsgrvpswl pvataeevat laddvltgat eapavavmea vadegsvlal
2941 tvrvldvvqc wlaggglegt klaivtrgav pagdgvvhdp aaaavwglvr aaqaenpdri
3001 vlldvepead vppllgsvla dgepqvavrg ttlsiprlar aarpdpaagf ktrgpvlvtg
3061 gtgslgglva rhlverhgvr qlvlasrrgl daegakdlvt dltalgadva vaacdvadrd
3121 qvaalltehr psavvhtagv pdagvigtvt pdrlaevfap kvtaarhlde ltrdldldsf
3181 vvyssvsavf mgagsgsyaa anayldglma hrraaglpgq slawglwdqt tggmaagtde
3241 agrarmtrrg glvamkpaag ldlfdaaigs gepllvpaql dlrglraeaa ggtevphllr
3301 glvragrqqa raastveenw agrlaglepa ergqvllelv raqvagvlgy raahqvdpdq
3361 glfeigfdsl taielrnrlr arterkispg vvfdhptpal laahlnellr kkv
```

RifKS10

RifAT10

RifDH10

RifKR10

RifACP10

Figure 3-3. Rifamycin PKS partial module 9 and module 10. Rifamycin polyketide synthase [*Amycolatopsis mediterranei* U32] Length: 3413

ADTASDAVSLGLAGADHPLLGAVVQLPQSDGLVFTSRLSLRSHPWLADHAVRDVVIVPGT
GLVELAVRAGDEAGCPVLDELVIEAPLVVPRRGGVRVQVALGGPADDGSRTVDVFSLRED
ADSWLRHATGVLVPENRPRGTAAFDFAAWPPPEAKPVDLTGAYDVLADVGYGYGPTFRAV
RAVWRRGSGNTTETFAEIALPEDARAEAGRFGIHPALLDAALHSTMVSAAADTESYGDEV
RLPFAWNGLRLHAAGASVLRVRVAKPERDSLSLEAVDESGGLVVTLDSLVGRPVSNDQLT
TAAG

ATGGCGGACACGGCAAGCGACGCAGTAAGCCTCGGCCTGGCAGGTGCGGATCACCCGCTGCTGGG
TGCAGTCGTGCAATTGCCGCAGAGCGATGGCCTGGTCTTTACGTCCCGCTTGAGCCTGCGCTCCC
ATCCGTGGTTGGCGGATCACGCGGTTCGCGACGTTGTTATTGTGCCGGGCACTGGTCTGGTTGAA
CTGGCCGTCCGTGCTGGCGACGAAGCCGGCTGCCCGGTGCTGGATGAGCTGGTGATCGAAGCGCC
GCTGGTCGTCCCGCGTCGTGGCGGTGTGCGTGTTCAAGTCGCACTGGGTGGTCCGGCGGATGAC
GGTTCGCGCACCGTTGACGTCTTTAGCCTGCGTGAGGACGCCGATAGCTGGCTGCGTCACGCTAC
CGGTGTGCTGGTTCCAGAGAATCGTCCGCGTGGTACCGCTGCCTTCGATTTTGCGGCGTGGCCTC
CGCCGGAAGCGAAGCCGGTCGACCTGACGGGCGCATACGATGTTTTGGCGGACGTTGGTTACGGT
TATGGCCCGACGTTCCGCGCAGTGCGTGCCGTGTGGCGTCGCGGCTCCGGTAACACCACCGAAAC
CTTTGCGGAGATCGCGCTGCCGGAGGACGCGCGTGCGGAGGCAGGCCGTTTCGGTATTCATCCG
GCACTGCTGGATGCCGCGCTGCATAGCACCATGGTCAGCGCCGCTGCGGATACCGAGAGCTATGG
CGATGAAGTTCGTCTGCCGTTCGCATGGAATGGTTTGCGCCTGCACGCGGCTGGTGCGAGCGTCC
TGCGTGTGCGCGTTGCCAAACCAGAACGCGATAGCCTGAGCCTGGAGGCGGTTGACGAGTCTGGT
GGCTTGGTTGTGACGCTGGACTCTCTGGTGGGTCGTCCTGTGAGCAACGACCAGCTGACCACTGC
AGCGGGT
N-terminus: CATATG (NdeI site)
C-terminus: TAACTCGAG (Stop codon/XhoI)

Figure 3-4. RifDH10 amino acid sequence. The synthetic gene encoding RifDH10 domain was subcloned in the pET-28a vector and the recombinant protein was expressed with a N-terminal His_6-tag in *E.coli* BL21(DE3). Protein expression and purification procedures are described below.

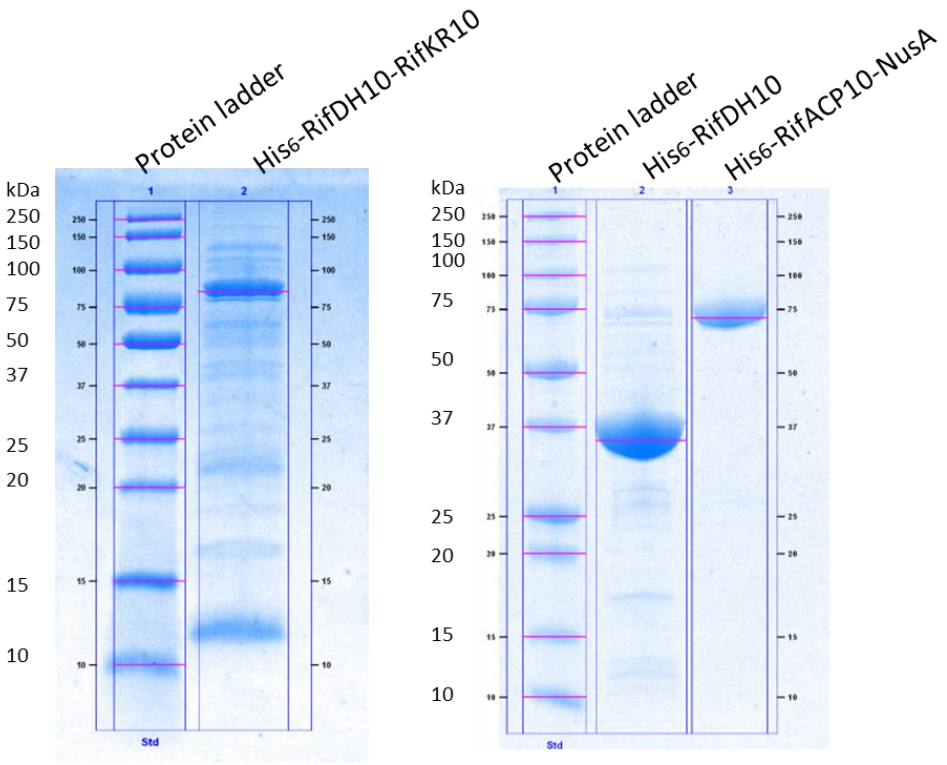


Figure 3-5. SDS-PAGE of recombinant RifDH10-KR10, RifDH10 and RifACP10NusA.

LAGLEPAERGQVLLELVRAQVAGVLGYRAAHQVDPDQGLFEIGFDSLTAIELRNRLRART
ERKISPGVVFDHPTPALLAAHLNELL

ATGCTGGCCGGCCTGGAACCTGCAGAACGCGGTCAAGTCCTGCTGGAGCTGGTGCGTGCGCAGGT
TGCGGGCGTGCTGGGCTACCGTGCTGCCCATCAGGTTGACCCGGATCAAGGTCTGTTCGAGATCG
GTTTTGACAGCCTGACGGCGATCGAGCTGCGCAACCGTTTGCGTGCGCGCACCGAACGTAAGATT
AGCCCGGGTGTCGTTTTCGATCACCCGACTCCGGCGTTGCTGGCAGCA
CACTTGAATGAGCTGCTG
N-terminus: CATATG (NdeI site)

Figure 3-6. RifACP10 amino acid sequence. The synthetic gene encoding RifACP10 was initially ligated into the NdeI/XhoI sites of pET-28a. Since the resultant recombinant RifACP10 was obtained only as insoluble inclusion bodies when expressed in *E.coli* BL21(DE3), the corresponding RifACP10-NusA protein was constructed.

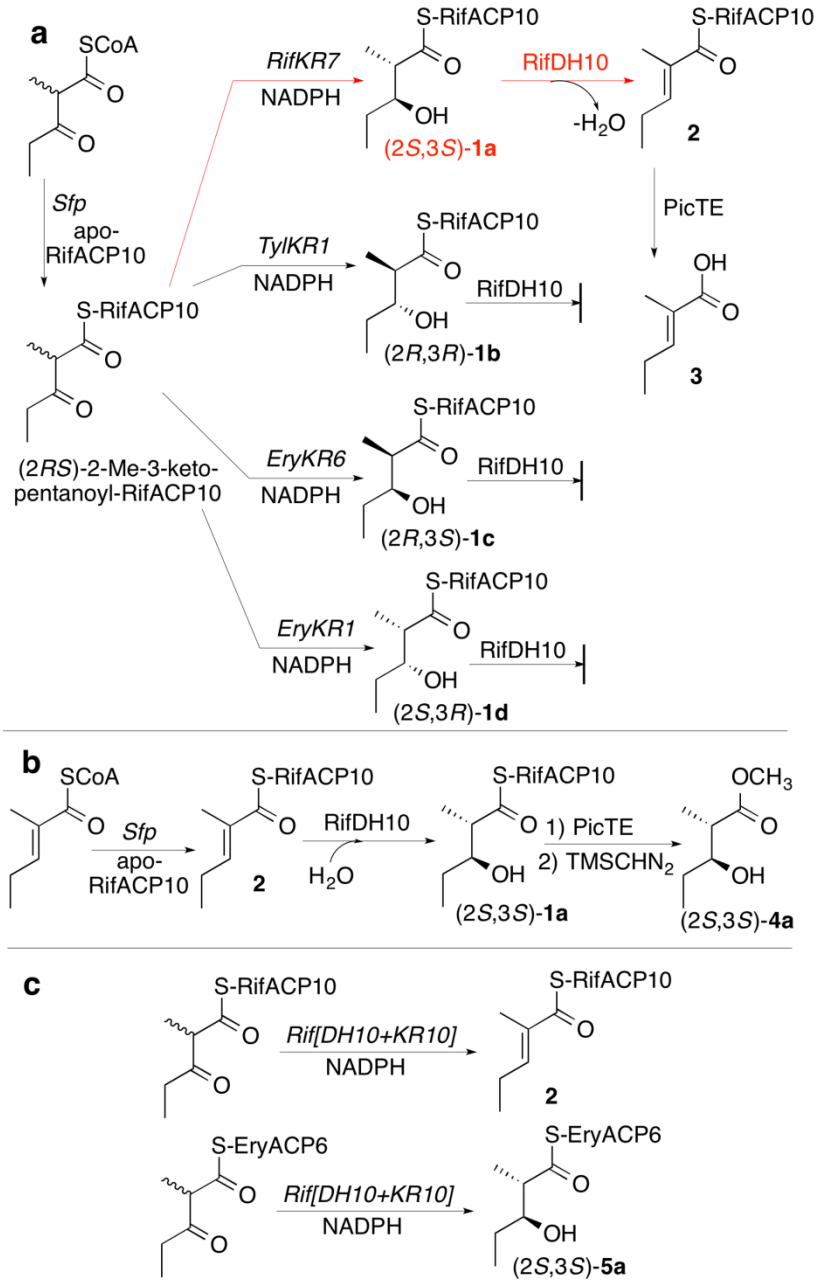


Figure 3-7. RifDH10-catalyzed dehydration/hydration of RifACP10-bound substrates. **(a)** Dehydration of (2*S*,3*S*)-2-methyl-3-hydroxypentanoyl-RifACP10 (1a) **(b)** Hydration of (*E*)-2-methyl-2-pentenoyl-RifACP10 (2). **(c)** Incubation of Rif[DH10+KR10] with RifACP10- and EryACP6-bound 2-methyl-3-ketoacyl thioesters.

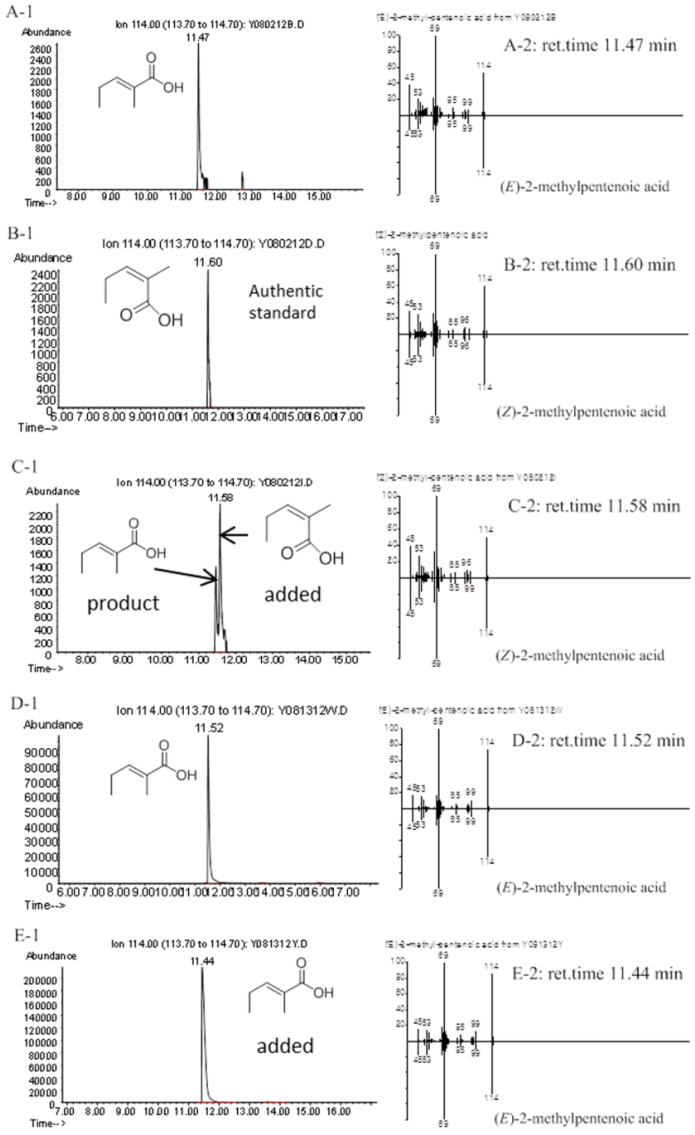


Figure 3-8. Chiral GC-MS analysis (Method 1) of the incubation of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA with RifKR7 and RifDH10. A and D: (*E*)-2-Methyl-2-pentenoic acid (3) from RifKR7-catalyzed reduction of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA followed by dehydration by RifDH10. B: (*Z*)-2-methylpentenoic acid authentic standard. C: Co-injection of A with B. E: Co-injection of (*E*)-2-methyl-2-pentenoic acid standard with D. A-1, B-1, C-1, D-1 and E-1: Extracted ion current (XIC) at *m/z* 114 (base peak). A-2, B-2, C-2, D-2 and E-2: mass spectra of selected peak, upper half, observed spectrum, lower half, inverted mass spectrum of reference standard.

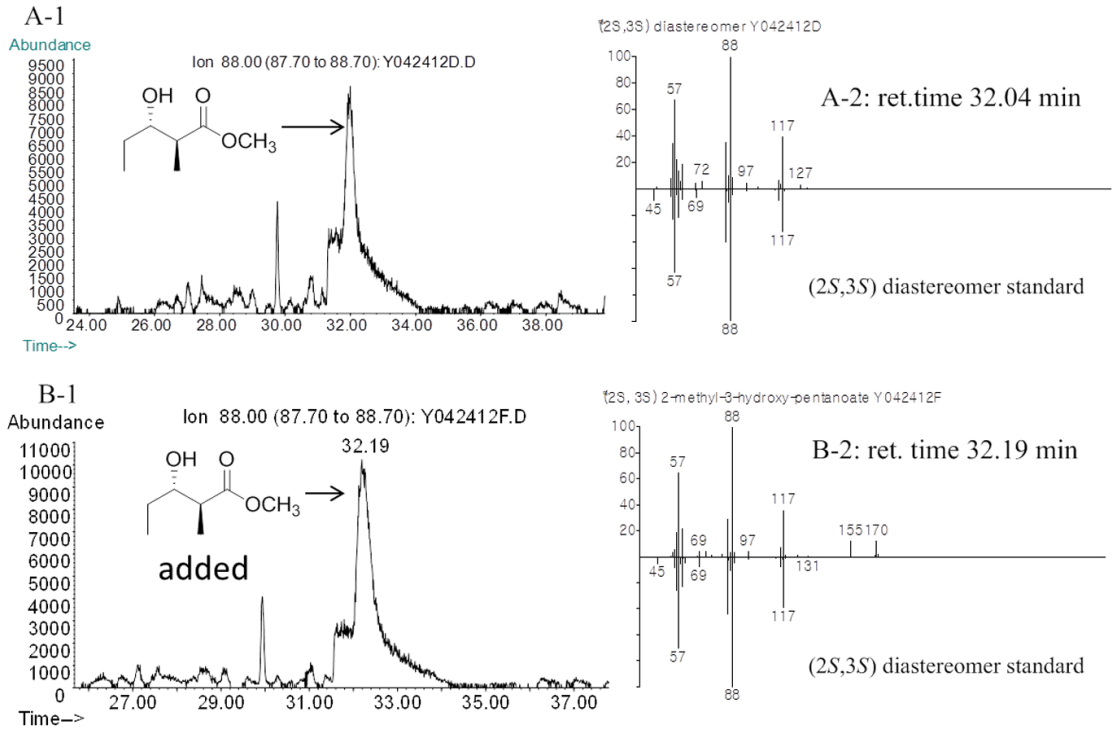


Figure 3-9. Chiral GC-MS analysis of the incubation of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA with RifKR7 in the absence of RifDH10 (Method 4). A: (2*S*,3*S*)-2-Methyl-3-hydroxypentanoate (4a) from RifKR7-catalyzed reduction of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA and B: Co-injection of (2*S*, 3*S*)-4a with A. A-1 and B-1: Extracted ion current (XIC) at *m/z* 88 (base peak). A-2 and B-2: Mass spectra of selected peak corresponding to (2*S*,3*S*)-4a upper half, observed spectra, lower half, inverted mass spectra of reference standard.

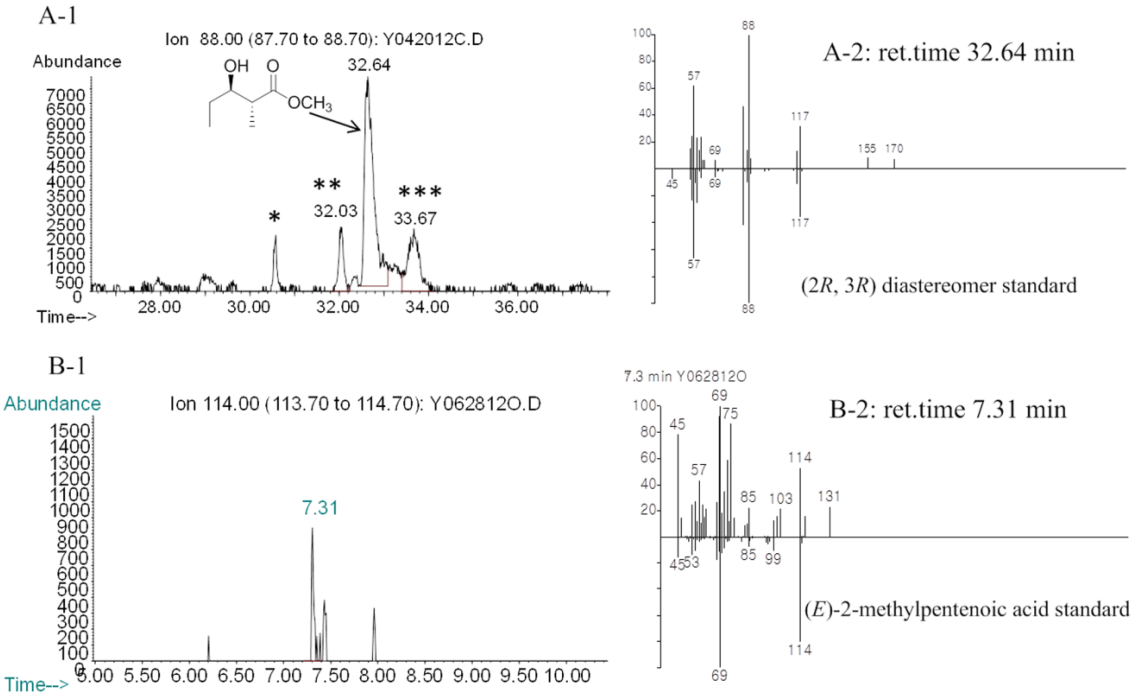


Figure 3-10. Chiral GC-MS analysis of the incubation of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA with TylKR1 and RifDH10 (Methods 2 and 4). A: 2-Methyl-3-hydroxypentanoates from TylKR1-catalyzed reduction of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA and B: Minor peak corresponding to (*E*)-2-methyl-2-pentenoic acid (3) generated by RifDH10-coupled dehydration of A. A-1: Extracted ion current (XIC) at *m/z* 88 (base peak). The major diastereomer observed was (2*R*,3*R*)-4b. The minor diastereomers are marked with asterisks as (2*R*,3*S*)-4c (*), (2*S*,3*R*)-4d (**) and (2*S*,3*S*)-4a (***). A-2: mass spectra of selected peak corresponding to (2*R*,3*R*)-4b, upper half, observed spectrum, lower half, inverted mass spectrum of reference standard. B-1: Extracted ion current (XIC) at *m/z* 114 (base peak). B-2: Mass spectrum of selected peak, upper half, observed spectrum, lower half, inverted mass spectra of reference standard. (Note both the lower intensity and lower purity of the ret. time 7.31 min peak containing 3 plus contaminants, compared to the yield and purity of 3 derived from an incubation carried out in the presence of RifKR7 (see Figure 3-8).

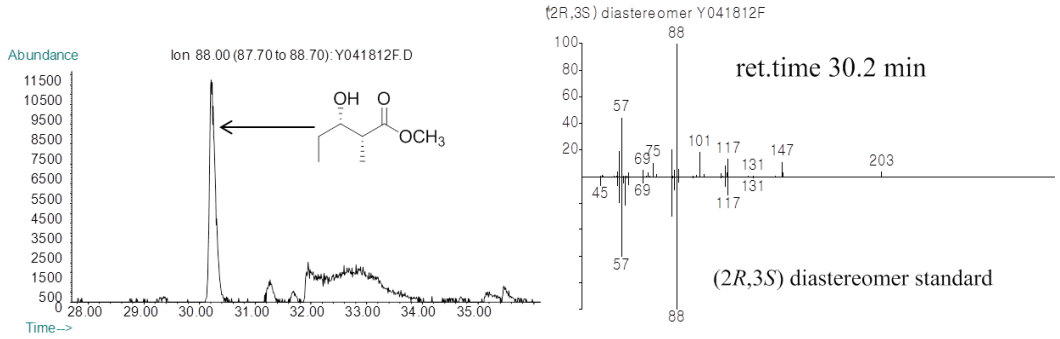


Figure 3-11. Chiral GC-MS analysis of the incubation of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA with EryKR6 and RifDH10 (Method 4). Left panel: Extracted ion current (XIC) at *m/z* 88 (base peak). Right panel, mass spectra of a selected peak corresponding to (2*R*,3*S*)-4c, upper half, observed spectrum, lower half, inverted mass spectrum of reference standard. No dehydration product 3 was observed.

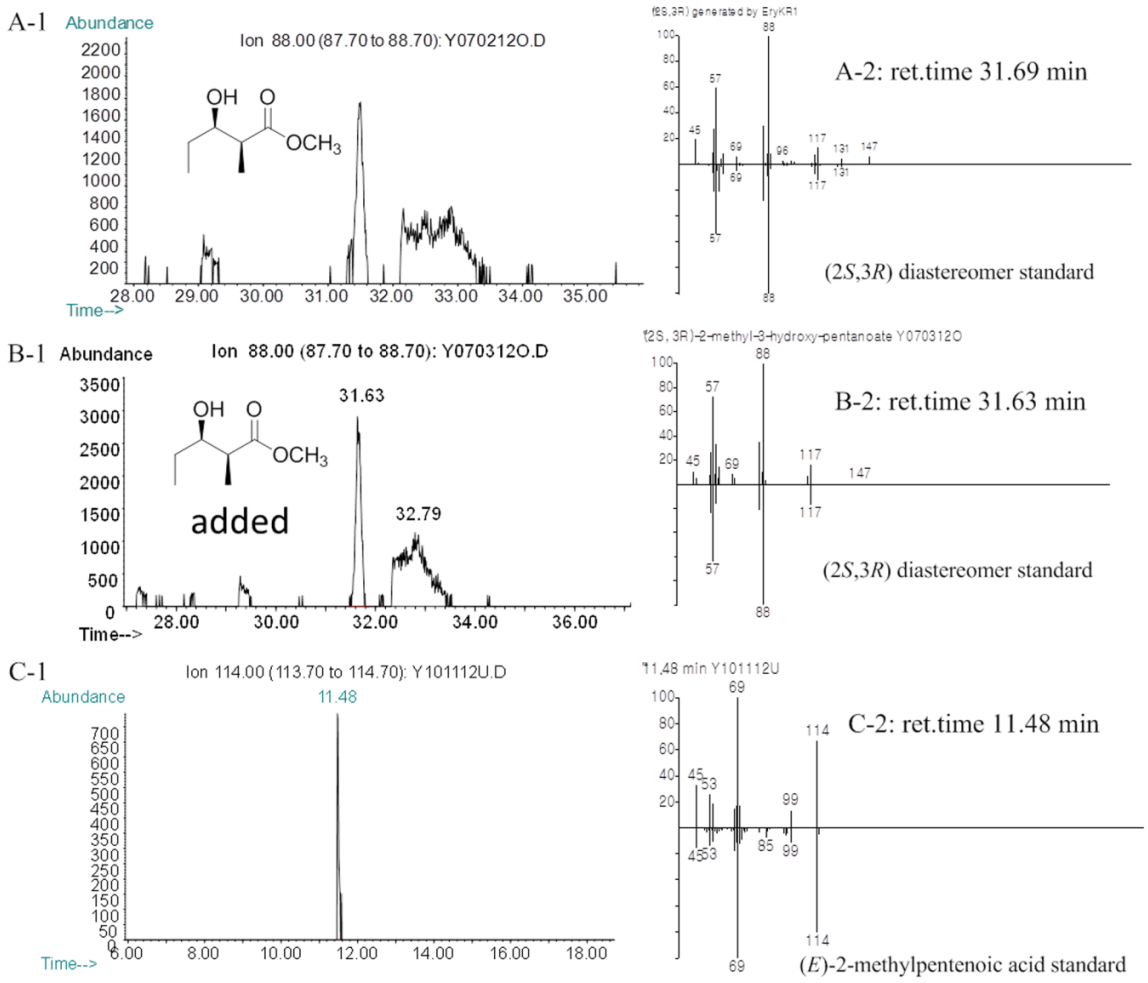


Figure 3-12. Chiral GC-MS analysis of the incubation of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA with EryKR1 and RifDH10. A: 2-Methyl-3-hydroxypentanoates from EryKR1-catalyzed reduction of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA (Method 4) and B: Co-injection of (2*S*,3*R*)-4d diastereomer with A. Envelope at 32.79 min contains minor amounts of (2*S*,3*S*)-4a. C: Minor amount of 3 generated by RifDH10-coupled reaction of A, due to presence of (2*S*,3*S*)-1a in reduced RifACP10-bound diketide. (Method 1) A-1 and B-1: Extracted ion current (XIC) at *m/z* 88 (base peak). A-2 and B-2: mass spectra of selected peak corresponding to (2*S*,3*R*)-4d, upper half, observed spectrum, lower half, inverted mass spectrum of reference standard. C-1: Extracted ion current (XIC) at *m/z* 114 (base peak). C-2: mass spectrum of selected peak, upper half, observed spectrum, lower half, inverted mass spectrum of reference standard

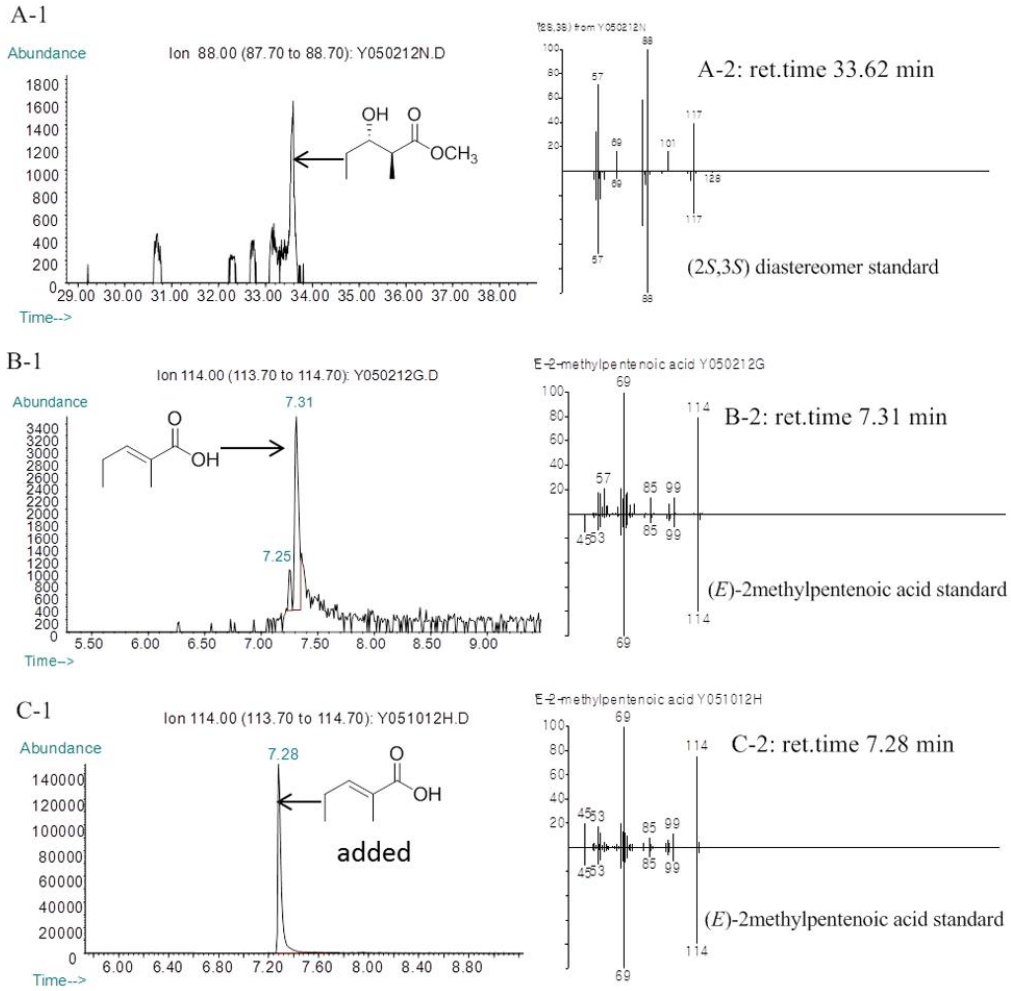


Figure 3-13. Chiral GC-MS analysis of the incubation of (2*RS*)-2-methyl-3-ketopentanoyl-RifACP10-NusA with RifDH10-KR10 (Methods 2 and 4). Left panels: A1. Extracted ion current (XIC) at *m/z* 88 (base peak for 4a) (Method 4); B-1 and C-1. Extracted ion current (XIC) at *m/z* 114 (base peak for 3) (Method 2). Right panels, mass spectra of selected peaks, upper half, observed spectrum, lower half, inverted mass spectrum of reference standard. A. (2*S*,3*S*)-4a produced by RifDH10-KR10. B. (*E*)-2-Methylpentenoic acid (3) produced by RifDH10-KR10. C. B plus authentic standard of (*E*)-3.

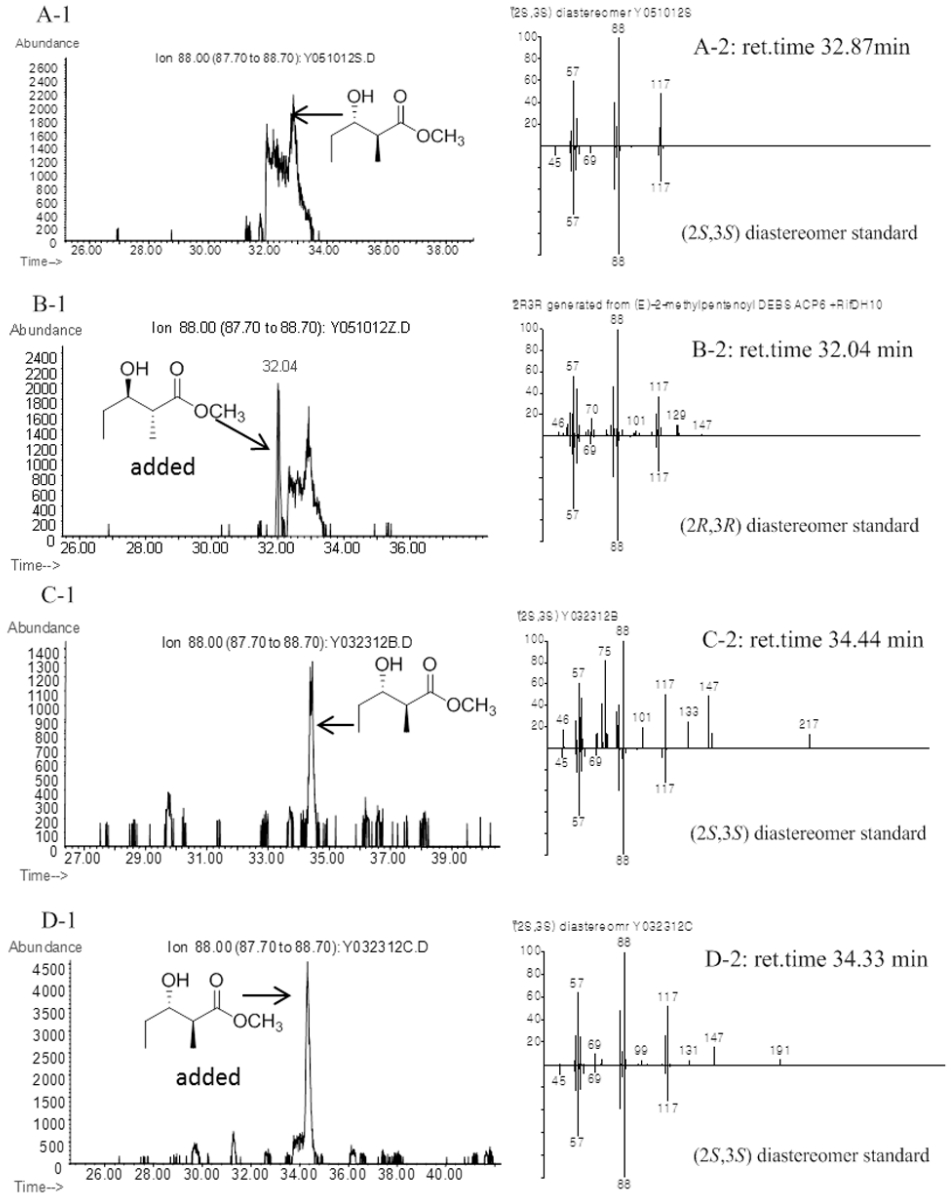


Figure 3-14. Chiral GC-MS analysis (Method 4) of the incubation of (*E*)-2-methyl-2-pentenoyl-RifACP10-NusA with RifDH10. Left panels: Extracted ion current (XIC) at *m/z* 88 (base peak). Right panels, mass spectra of selected peaks corresponding to diastereomers of methyl 2-methyl-3-hydroxypentanoate, upper half, observed spectrum, lower half, inverted mass spectrum of reference standard. A and C. (2*S*,3*S*)-4a generated by RifDH10. B. A plus (2*R*,3*R*)-4b. D. C plus (2*S*,3*S*)-4a.

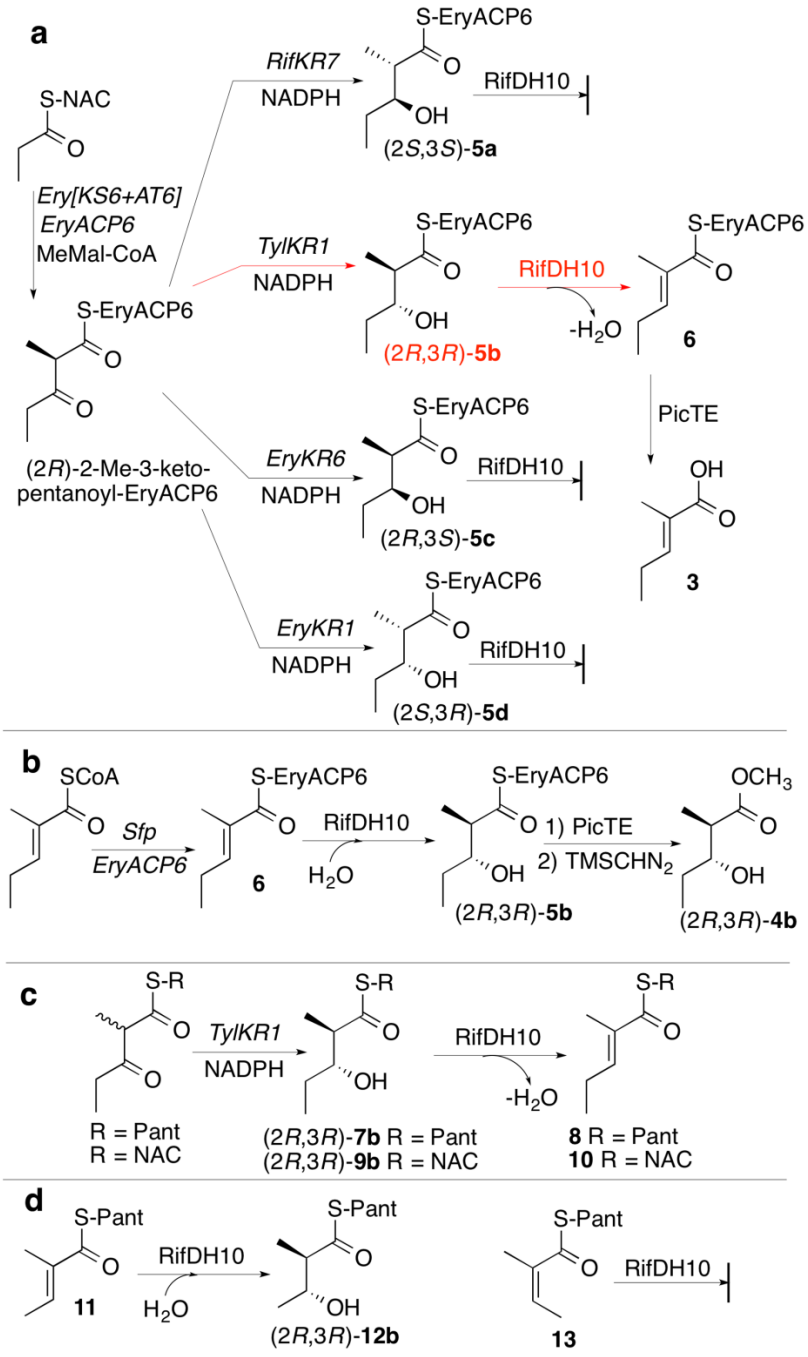
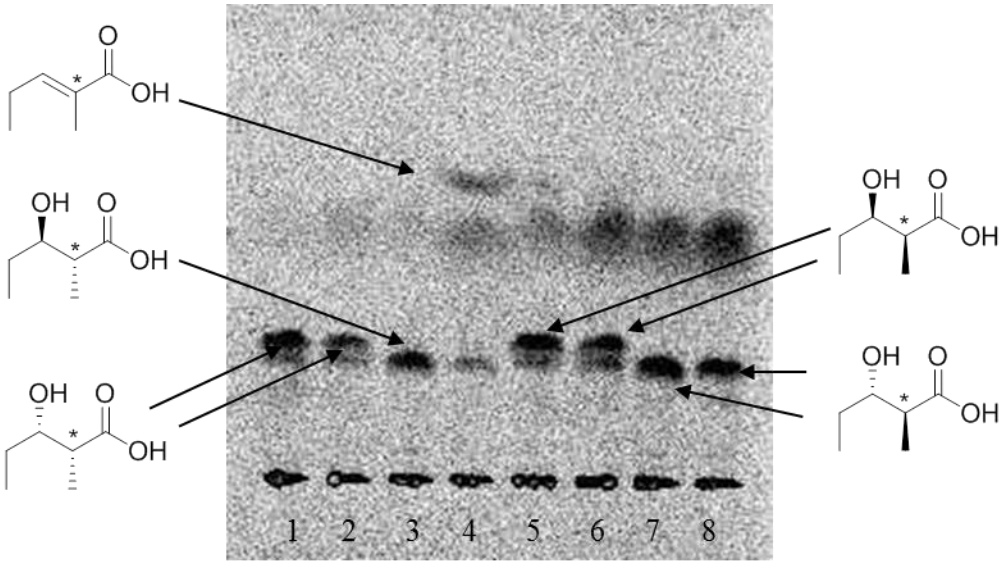


Figure 3-15. RifDH10-catalyzed dehydration/hydration of acyl thioester analogues. **(a)** Dehydration of (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-EryACP6 (5b). **(b)** Hydration of (*E*)-2-methyl-2-pentenoyl-EryACP6 (6). **(c)** Dehydration of (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-*S*-Pant (7b) and –S-NAC (9b) analogues. **(d)** Incubation of RifDH10 with (*E*)-11 and (*Z*)-13.



	EryKR6	TylKR1	EryKR1	RifKR7	RifDH10
1	+	-	-	-	-
2	+	-	-	-	+
3	-	+	-	-	-
4	-	+	-	-	+
5	-	-	+	-	-
6	-	-	+	-	+
7	-	-	-	+	-
8	-	-	-	+	+

Figure 3-16. TLC-phosphorimaging of diketide acid products from combinatorial enzyme reactions involving propionyl-SNAC, Ery[KS6][AT6], EryACP6, RifDH10, [^{14}C]-methylmalonyl-CoA, NADPH, and varying KR domains. Only Expt 4 with TylKR1 and RifDH10 produced 2-methyl-2-pentenoic acid. Note that the achiral TLC assay cannot distinguish *syn*-(2*S*,3*R*)- from the enantiomeric *syn*-(2*R*,3*S*)-2-methyl-3-hydroxypentanoic acid nor *anti*-(2*S*,3*S*)- from the enantiomeric *anti*-(2*R*,3*R*)-2-methyl-3-hydroxypentanoic acid.

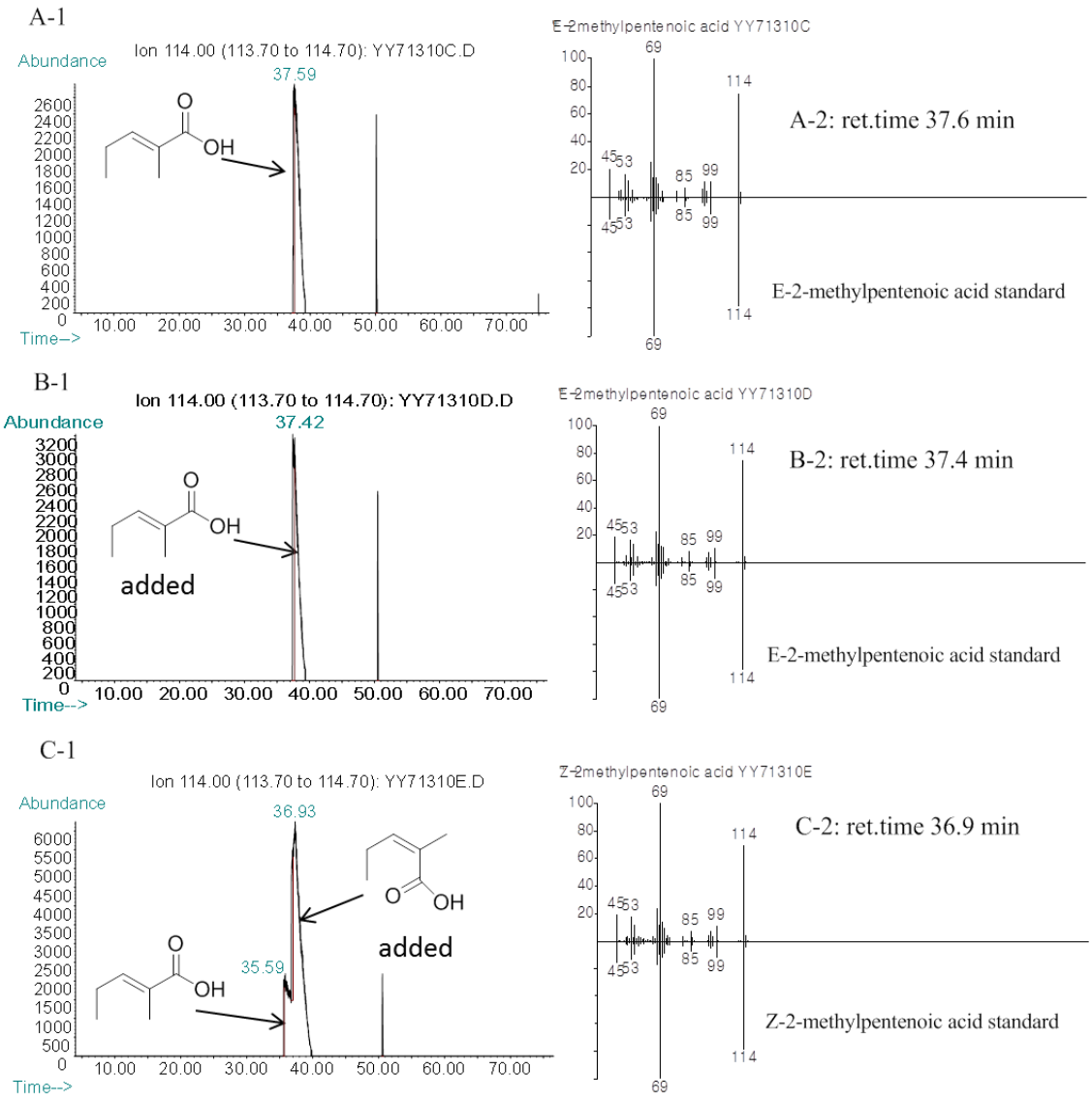


Figure 3-17. GC-MS analysis of the incubation of in situ-generated (2*R*)-2-methyl-3-ketopentanoyl-EryACP6 with TylKR1, NADPH, and RifDH10 (Method 3). Left panels: Extracted ion current (XIC) at *m/z* 114 (base peak). Right panels, mass spectra of selected peaks corresponding to unsaturated diketide acid 3, upper half, observed peaks, lower half, mass spectrum of reference standard 3. A. Reaction product. B. A plus authentic (*E*)-3. C. A plus authentic (*Z*)-2-methyl-2-pentenoic acid.

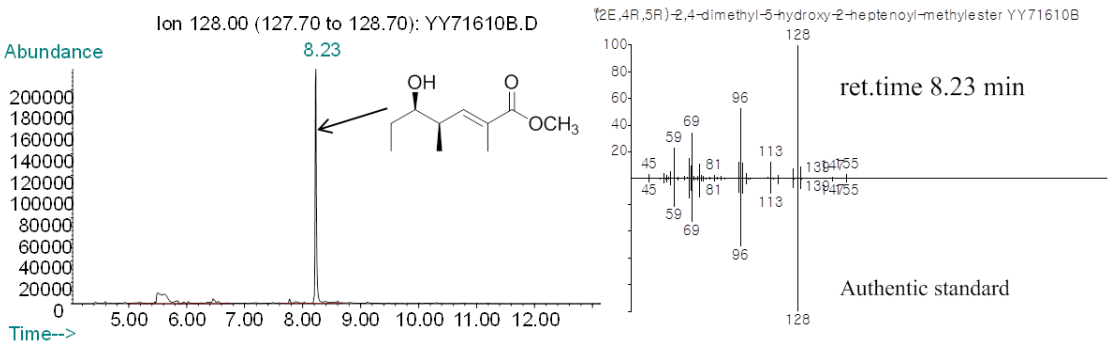


Figure 3-18. Chiral GC-MS analysis of the incubation of in situ-generated (2*R*,4*S*,5*R*)-2,4-dimethyl-3-keto-5-hydroxyheptanoyl-EryACP6 with TylKR1 and RifDH10. Reaction produces (2*E,4R,5R*)-2,4-dimethyl-5-hydroxy-2-heptenoyl methyl ester resulting from dehydration of the intermediate (2*R*,3*R,4S,5R*)-2,4-dimethyl-3,5-dihydroxyheptanoyl-EryACP6, hydrolysis, and methylation. Left panel: Extracted ion current (XIC) at *m/z* 128 (base peak for (2*E,4R,5R*)-2,4-dimethyl-5-hydroxy-2-heptenoyl methyl ester). Right panel: mass spectrum of selected peak corresponding to the unsaturated triketide acid methyl ester, upper half, observed peak, lower half, mass spectrum of reference standard.

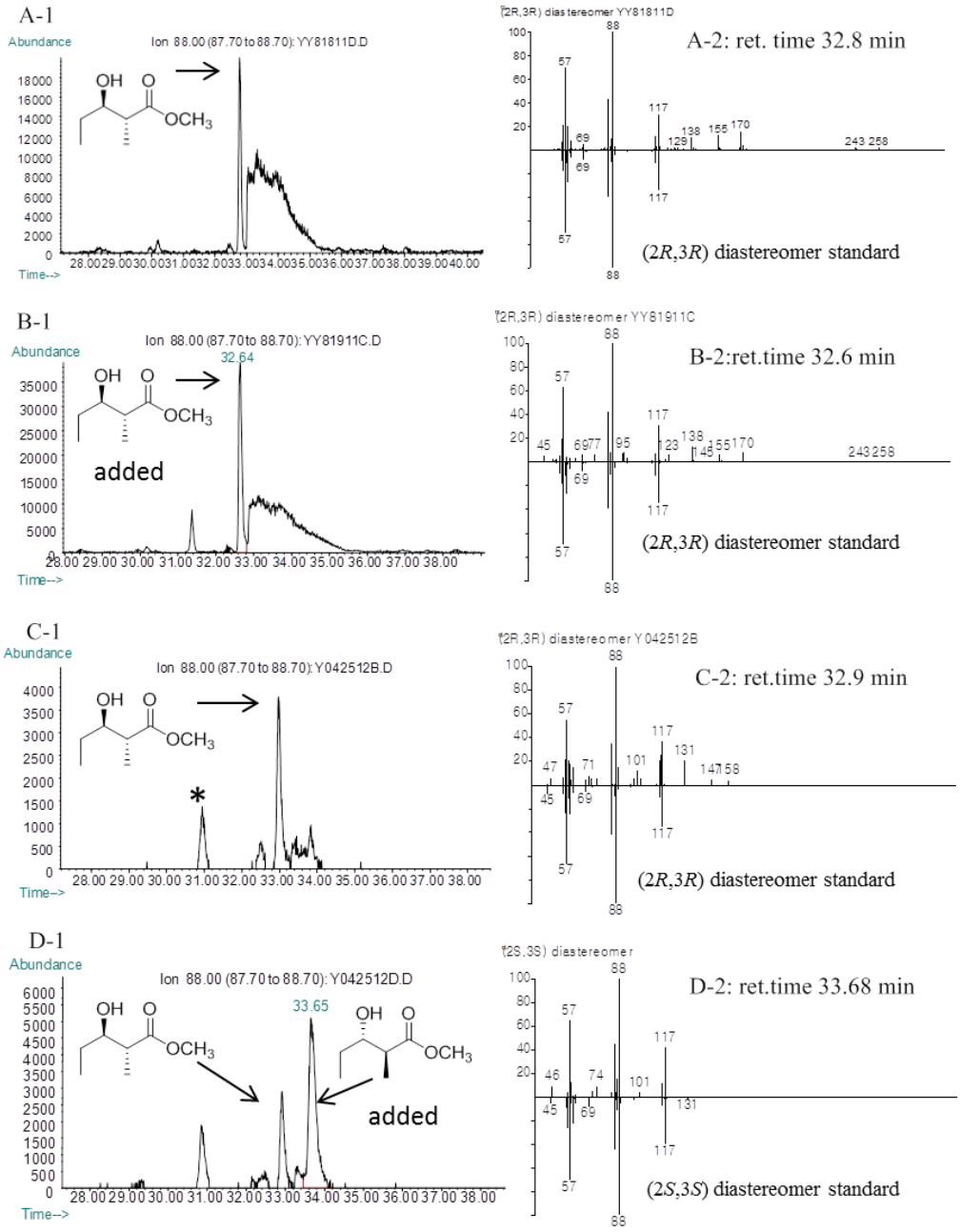


Figure 3-19. Chiral GC-MS analysis of the incubation of (*E*)-2-methylpentenoyl-EryACP6 with RifDH10 (Method 4). Left panels: Extracted ion current (XIC) at *m/z* 88 (base peak). Right panels, mass spectra of selected peaks corresponding to diastereomers, upper half, observed spectrum, lower half, inverted mass spectrum of reference standard. A and C. (2*R*,3*R)*-4b produced by RifDH10. B. A plus (2*R*,3*R*)-4b standard. D. C plus (2*S*,3*S*)-4a standard.

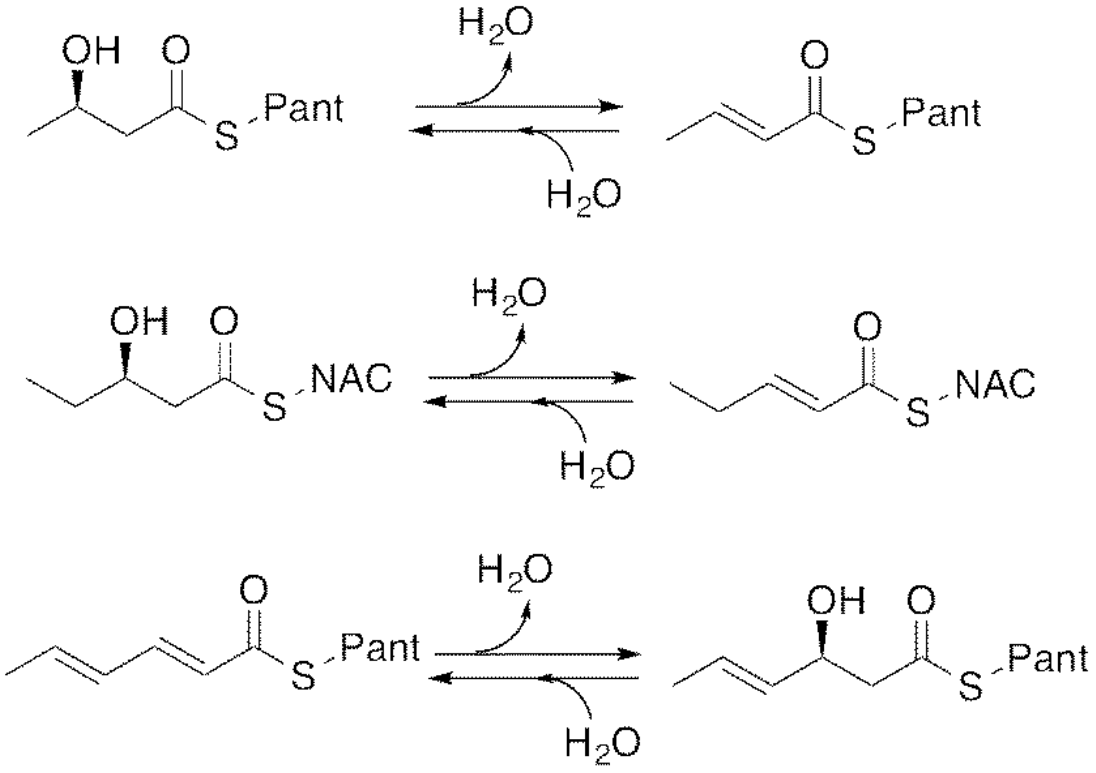


Figure 3-20. RifDH10-catalyzed dehydration/hydration of acyl thioester analogues.

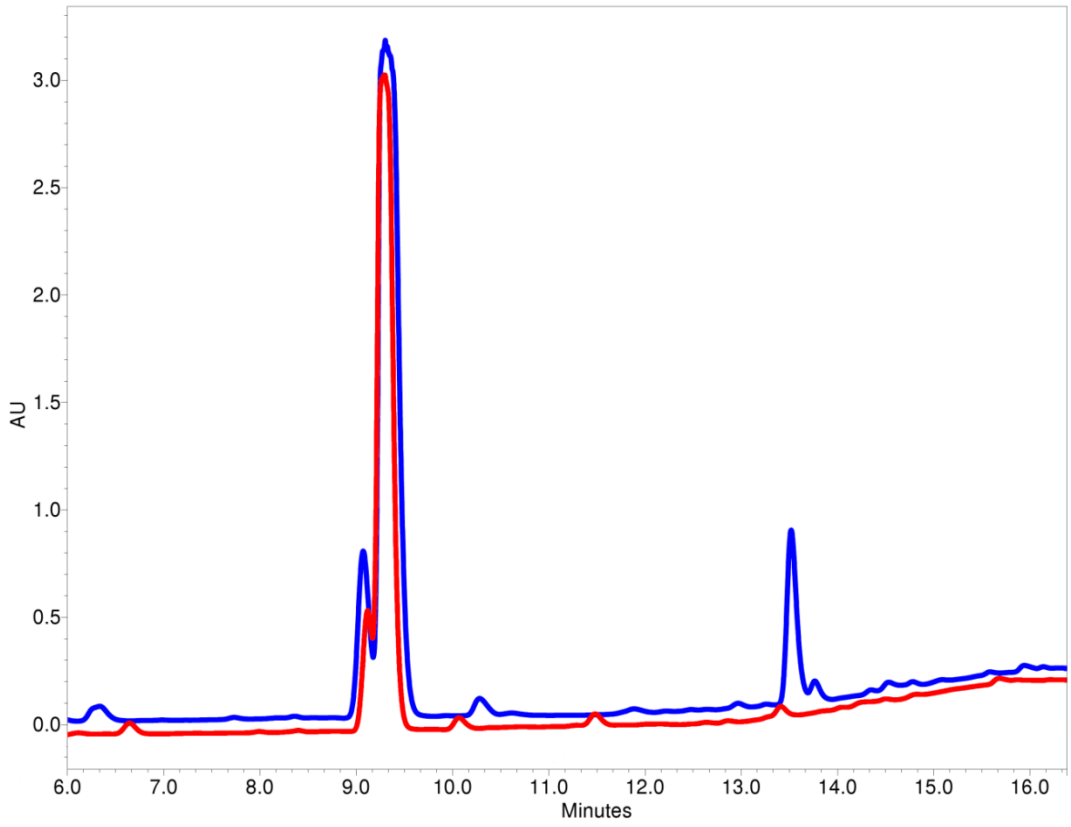


Figure 3-21. Dehydration of (3*R*)-3-hydroxybutanoyl-*S*-NAC. Reversed-phase HPLC analysis of a control (red, no enzyme) and the RifDH10-catalyzed dehydration (blue) of (3*R*)-3-hydroxybutanoyl-*S*-NAC (9.2 min), with product eluting at 13.5 min. Product peak was collected and confirmed with LC/MS: expected mass 187.3; observed mass: 187.0.

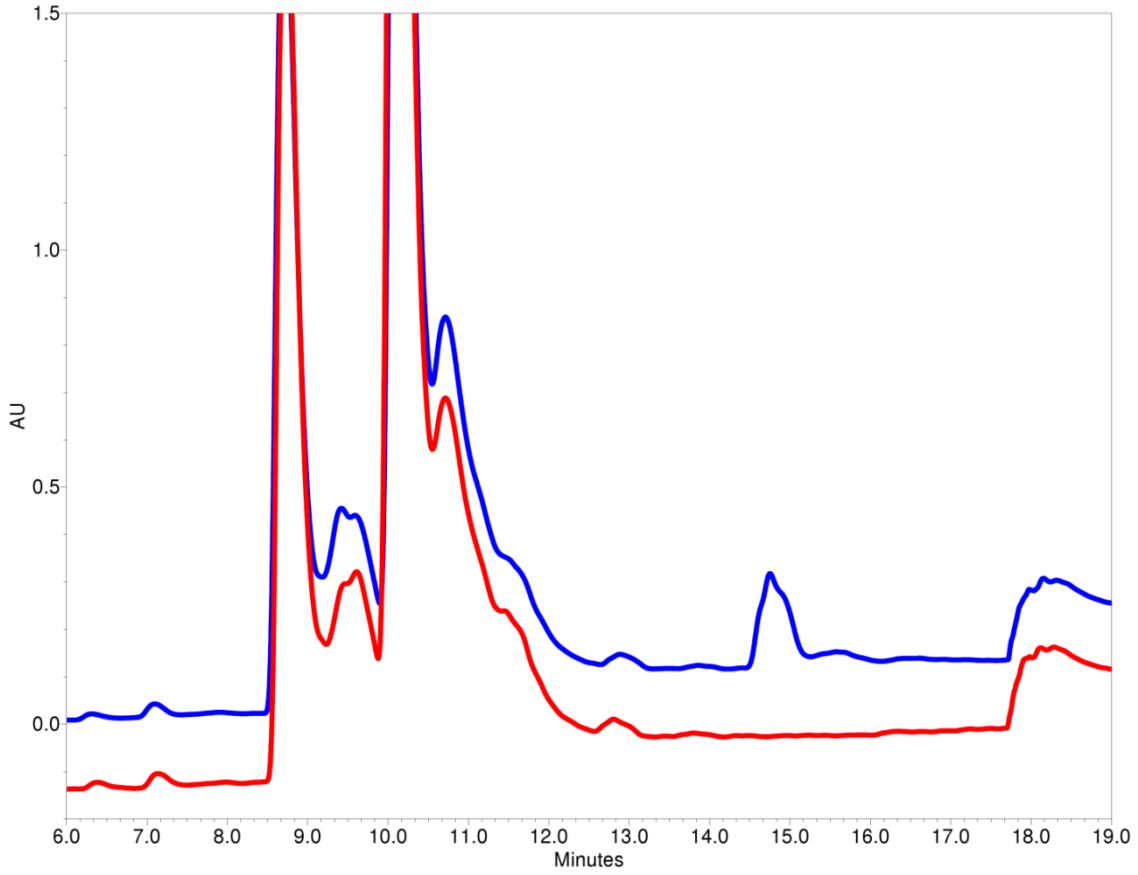


Figure 3-22. Dehydration of (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-*S*-NAC. Reversed-phase HPLC analysis of a control (red, no enzyme) and the RifDH10 catalyzed dehydration (blue) of (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-*S*-NAC (10.2 min), with product eluting at 14.8 min. The significant peak at 8.9 min is residual (2*RS*)-2-methyl-3-oxopentanoyl-*S*-NAC which was not reduced by the preceding TylKR1 catalyzed reduction. Product peak was collected and confirmed with LC/MS: expected mass 215.3; observed mass: 216.0.

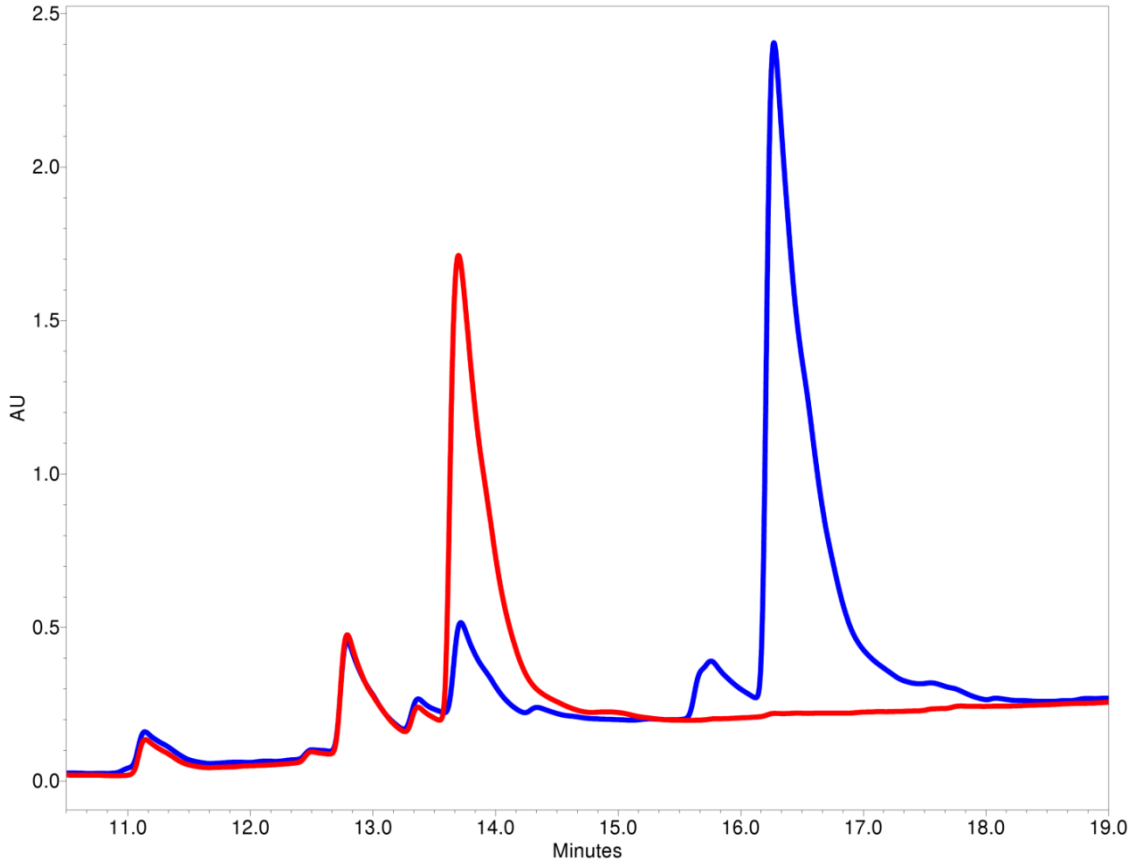


Figure 3-23. Dehydration of (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-*S*-pantetheine. Reversed-phase HPLC analysis of a control (red, no enzyme) and the RifDH10-catalyzed dehydration (blue) of (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl-*S*-pantetheine (13.8 min), with product eluting at 16.4 min. Product peak was collected and confirmed with LC/MS: expected mass 374.5; observed mass: 374.0.

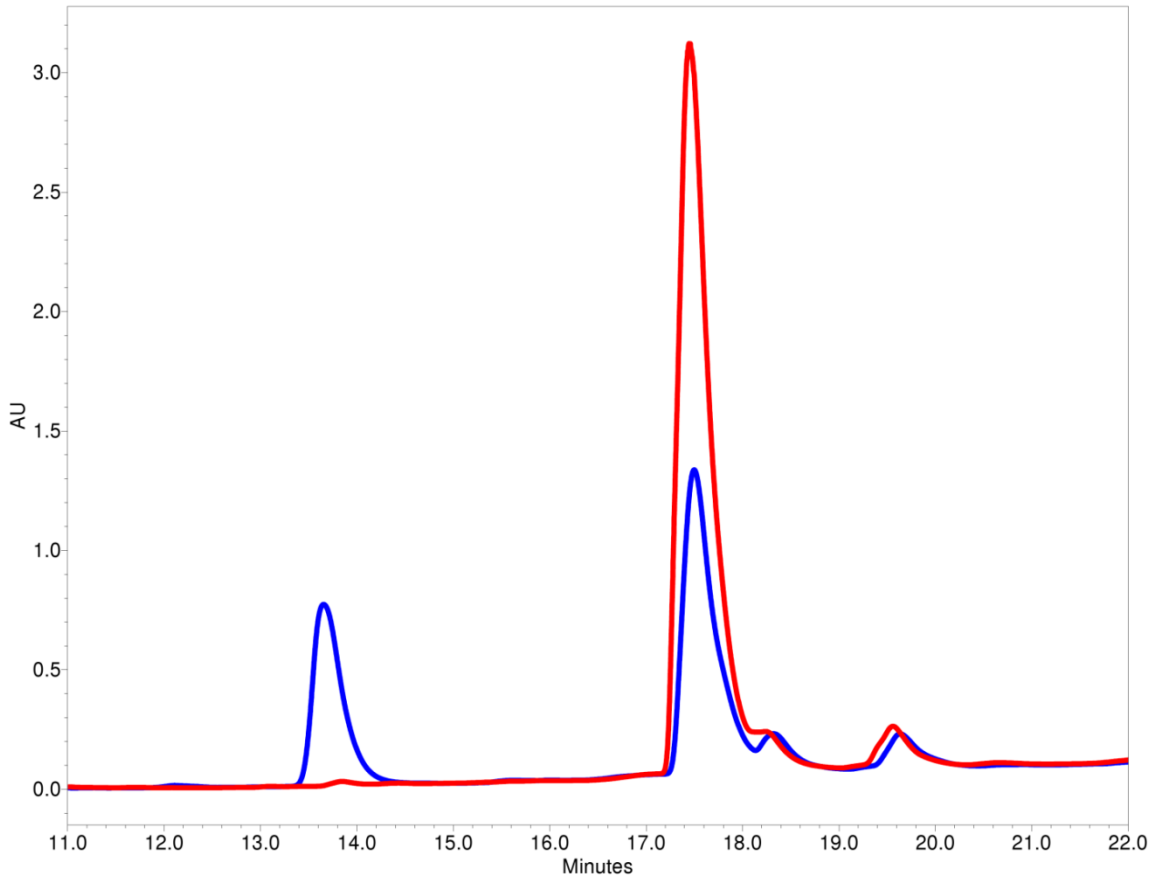


Figure 3-24. Hydration of (*E*)-2-butenoyl-*S*-pantetheine. Reversed-phase HPLC analysis of a control (red, no enzyme) and the RifDH10-catalyzed hydration (blue) of (*E*)-2-butenoyl-*S*-pantetheine (17.6 min), with product eluting at 13.8 min. Product peak was collected and confirmed with LC/MS: expected mass 364.5; observed mass: 365.0.

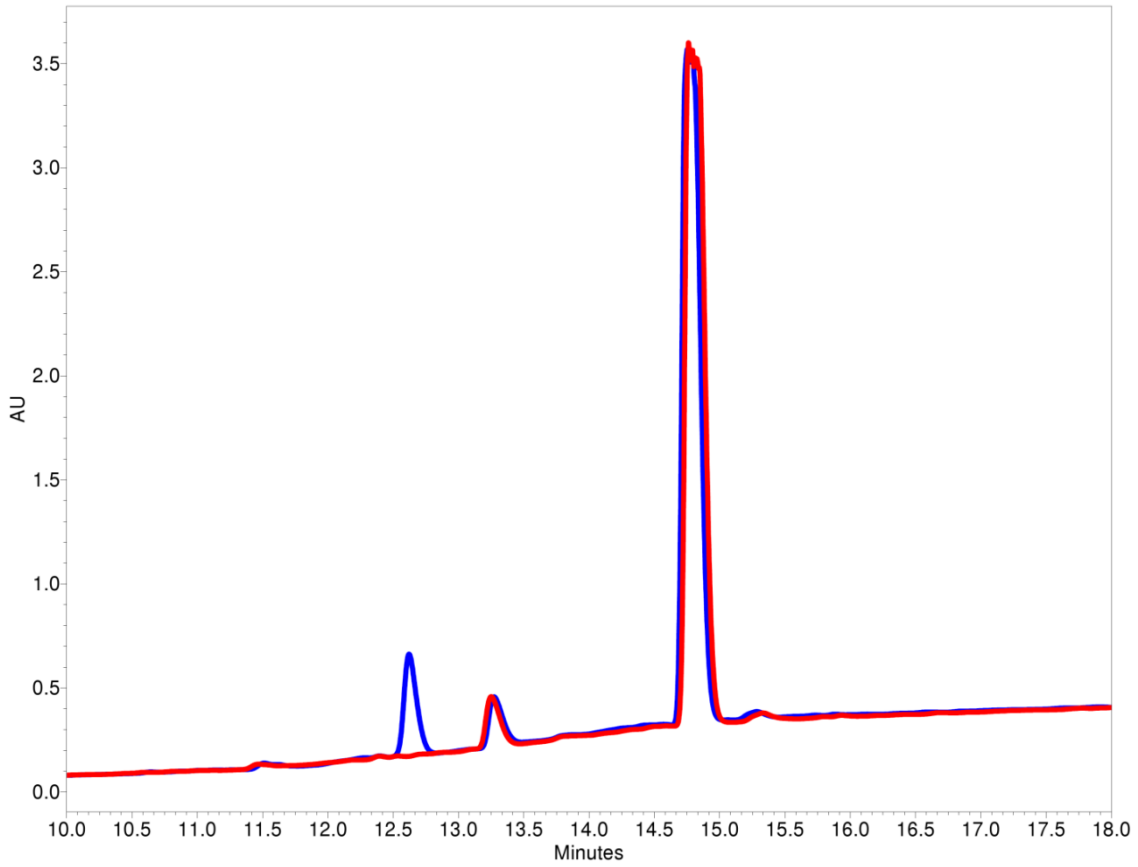


Figure 3-25. Hydration of (*E*)-2-methyl-2-butenoyl-*S*-pantetheine. Reversed-phase HPLC analysis of a control (red, no enzyme) and the RifDH10-catalyzed hydration (blue) of (*E*)-2-methyl-2-butenoyl-*S*-pantetheine (14.8 min), with product eluting at 12.6 min. Product peak was collected and confirmed with LC/MS: expected mass 378.5; observed mass: 378.0.

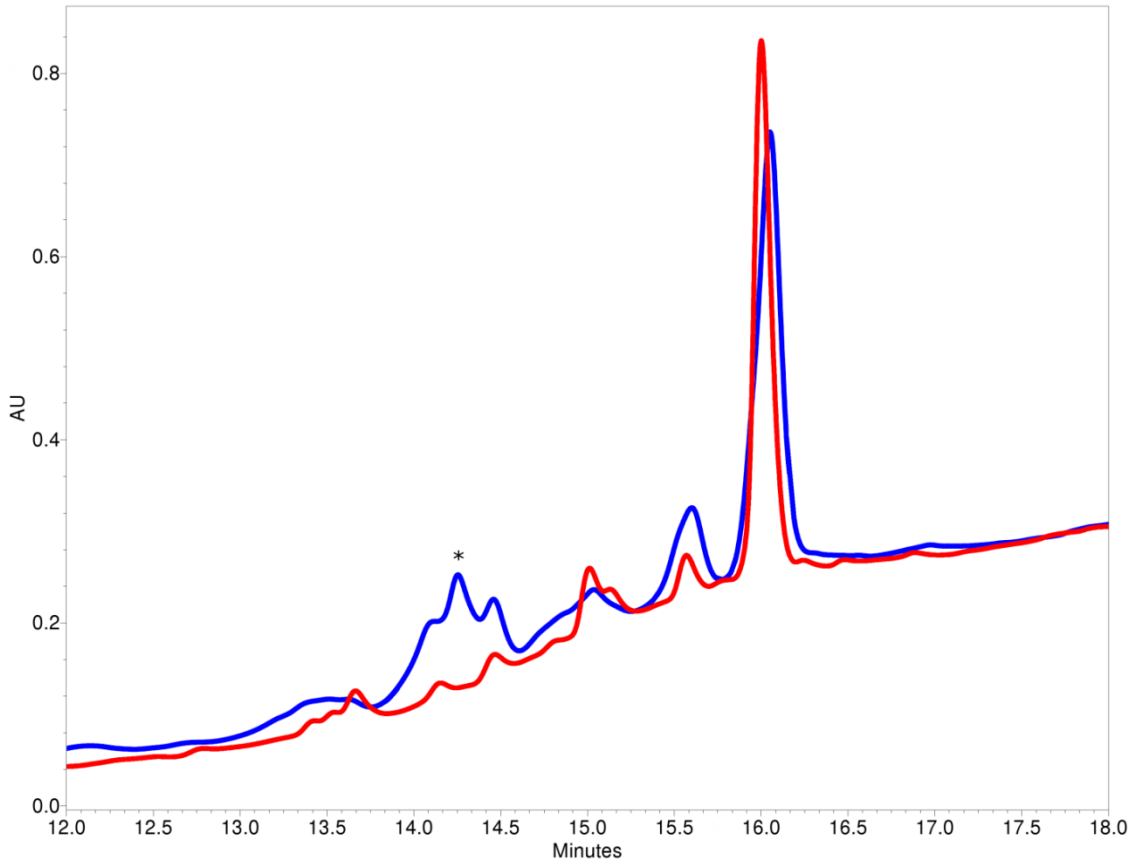


Figure 3-26. Hydration of (*E*,*E*)-2,4-hexadienoyl-*S*-pantetheine. Reversed-phase HPLC analysis of a control (red, no enzyme) and the RifDH10 catalyzed hydration (blue) of (*E*,*E*)-2,4-hexadienoyl-*S*-pantetheine (16.1 min), with product eluting at 14.3 min. The hydrated product has been highlighted with an asterisk due to adjacent contaminating peaks, which are also present in the control reaction. Product peak was collected and confirmed with LC/MS: expected mass 390.5; observed mass: 391.0.

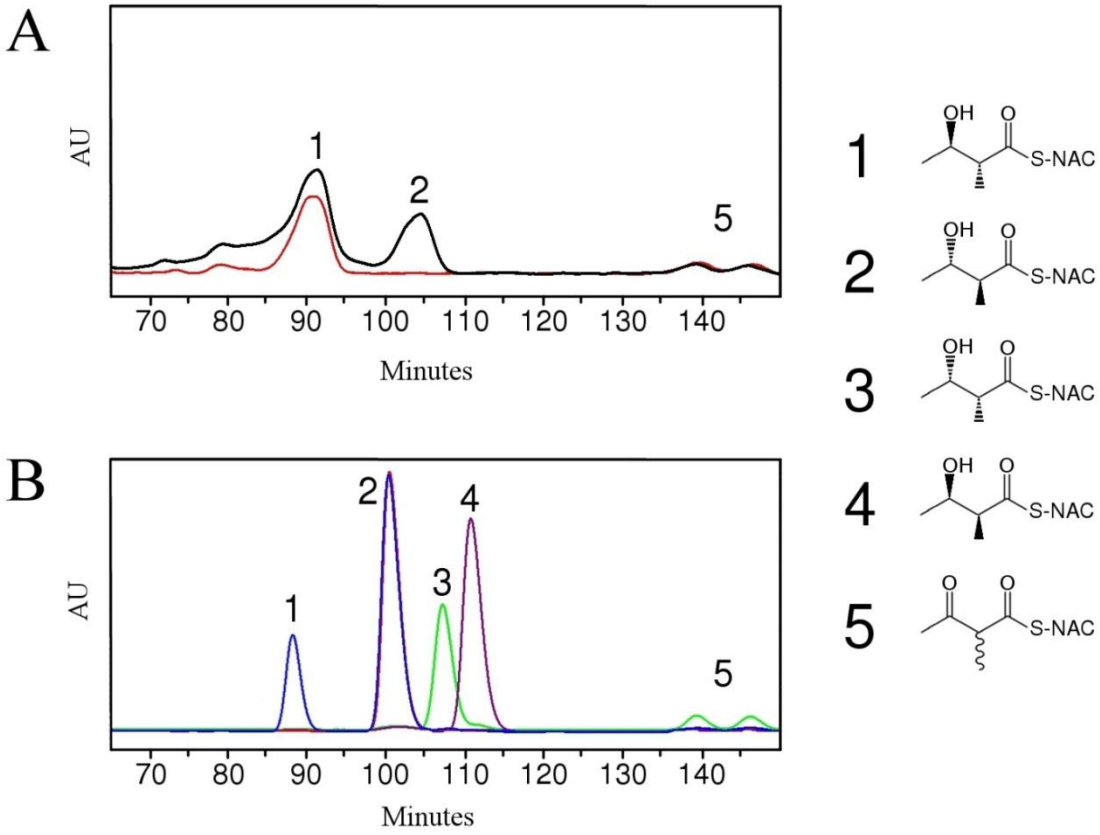


Figure 3-27. Stereochemistry of hydrated product from RifDH10-catalyzed hydration of (*E*)-2-methyl-2-butenoyl-*S*-pantetheine. (*E*)-2-Methyl-2-butenoyl-*S*-pantetheine was incubated with RifDH10, and the resultant hydrated product was purified by reverse-phase HPLC. The acyl group was subsequently transferred to *S*-NAC, and subjected to chiral HPLC analysis. **(a)** Chiral HPLC trace of the product of the RifDH10 catalyzed hydration of (*E*)-2-methyl-2-butenoyl-*S*-pantetheine (red). The run was repeated after being spiked with authentic (2*S*,3*S*)-2-methyl-3-hydroxy-butanoyl-*S*-NAC (black trace). **(b)** A series of authentic standards analyzed on the same chiral column used to determine retention times of each of the stereoisomers of 2-methyl-3-hydroxy-butanoyl-*S*-NAC.

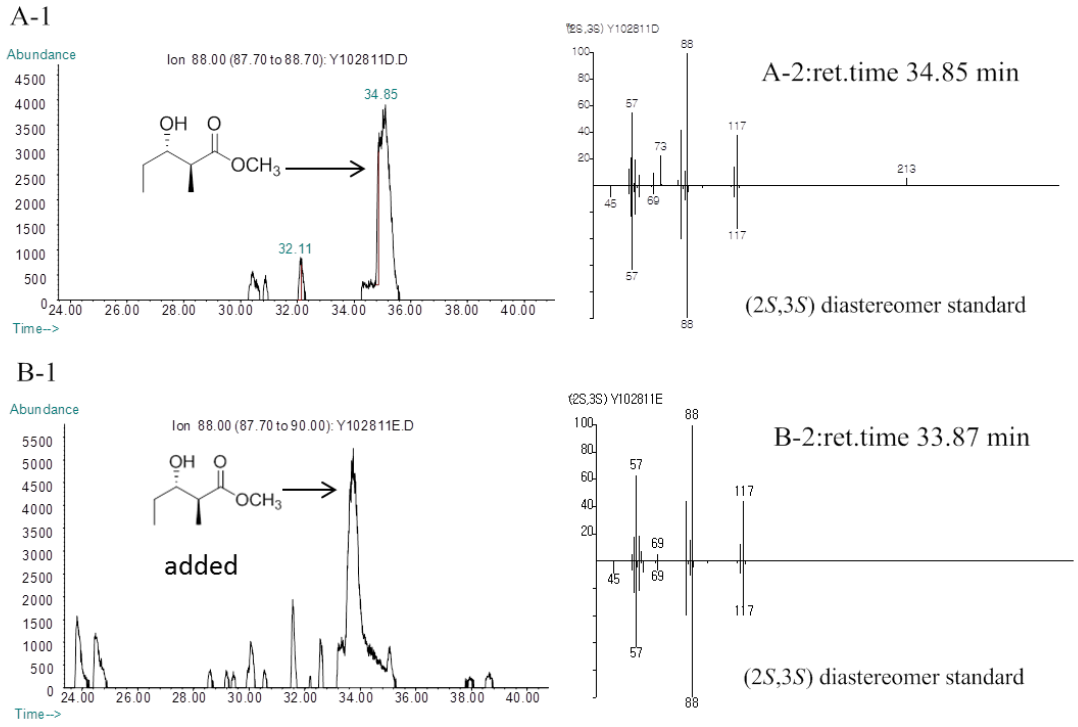


Figure 3-28. Chiral GC-MS analysis of the incubation of in situ-generated (2*R*)-2-methyl-3-ketopentanoyl-EryACP6 with RifDH10-KR10 (Method 4). Formation of methyl (2*S*,3*S*)-2-methyl-3-hydroxypentanoate. Left panels: Extracted ion current (XIC) at *m/z* 88 (base peak). Right panels, mass spectra of selected peaks corresponding to 4a, upper half, observed spectrum, lower half, inverted mass spectrum of reference standards. A (2*S*,3*S)*-4a generated by RifDH10-KR10. B. A plus (2*S*,3*S*)-4a generated by enzymatic incubation by RifKR7 alone. No dehydration products such as 3 could be detected.

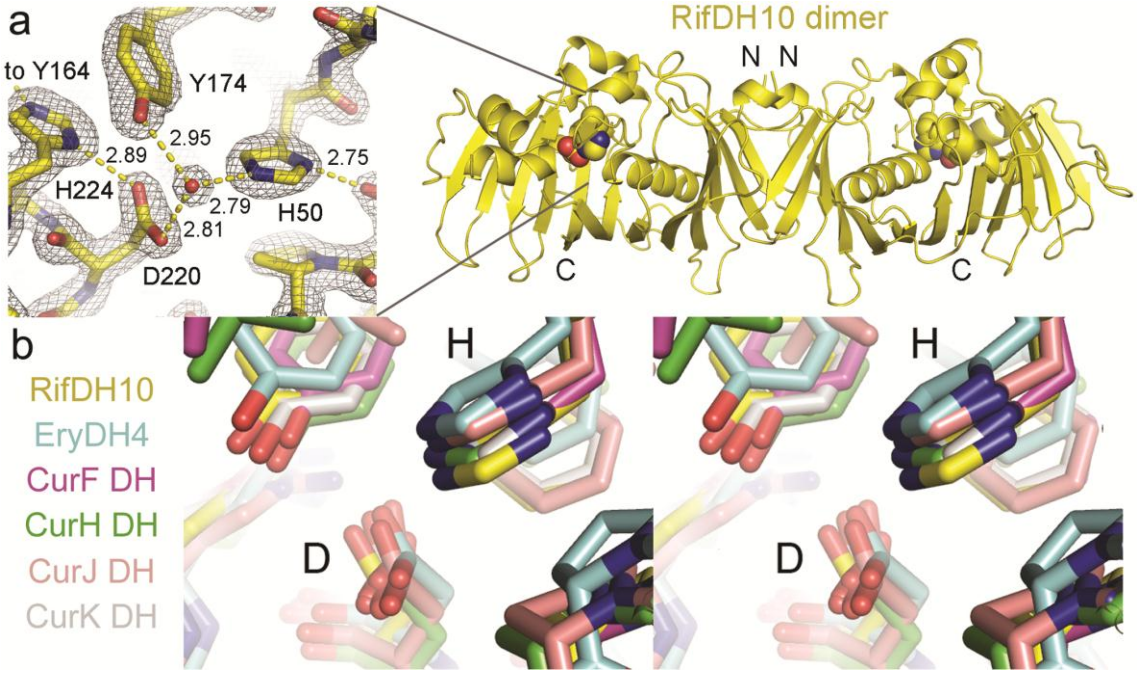


Figure 3-29. The 1.82 Å-resolution structure of RifDH10. **(a)** The $2F_o$-F_c electron density map (contoured at 1.8 σ) shows a water molecule, bound to the catalytic histidine (H50) and aspartate (D220), as well as a nearly invariant Y174, that is representative of the water molecule eliminated through a *syn*-dehydration reaction. H224 and Y164 may help increase the pK_a of D220 through a hydrogen-bond network. **(b)** Stereodiagram of the superposition of the active sites of RifDH10, EryDH4, CurF DH, CurH DH, CurJ DH, and CurK DH, showing the catalytic His (H) and Asp (D) residues.

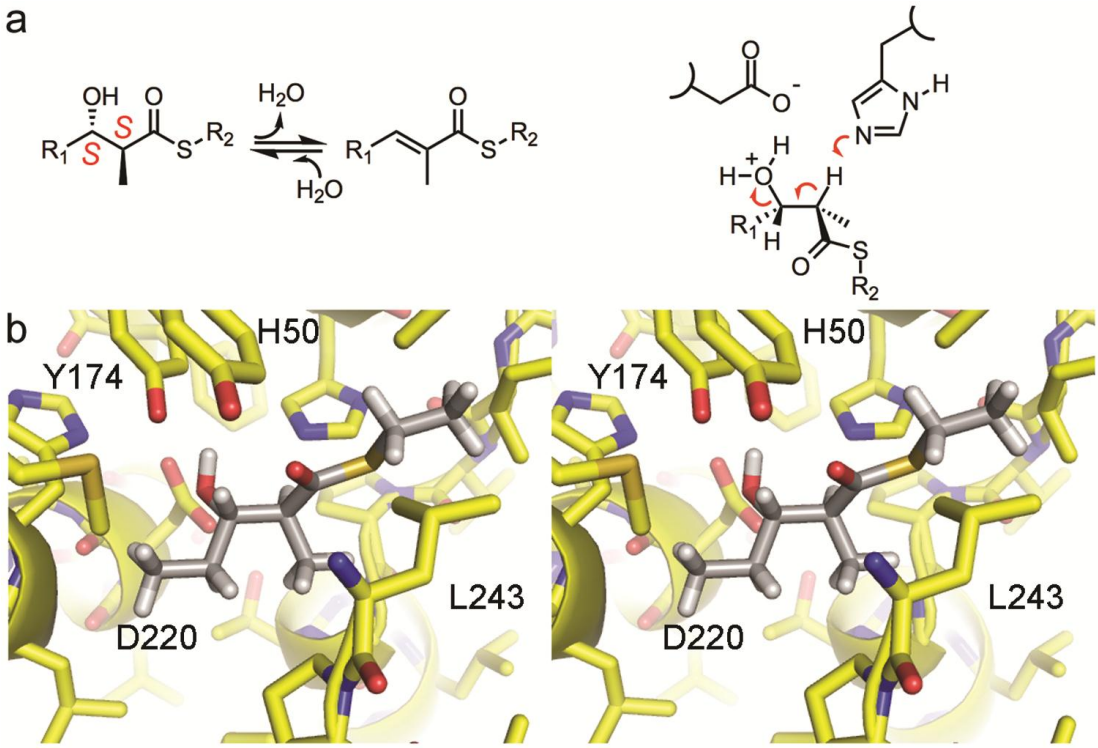


Figure 3-30. RifDH10-catalyzed dehydration. **(a)** *syn*-Dehydration of a (2*S*,3*S*)-2-methyl-3-hydroxyacyl thioester substrate generates a *trans* (*E*) double bond. **(b)** Stereodiagram showing (2*S*,3*S*)-2-methyl-3-hydroxyacyl thioester substrate modeled into the active site of RifDH10. See Figure 3-32 for modeling of diastereomeric thioesters into the active site.

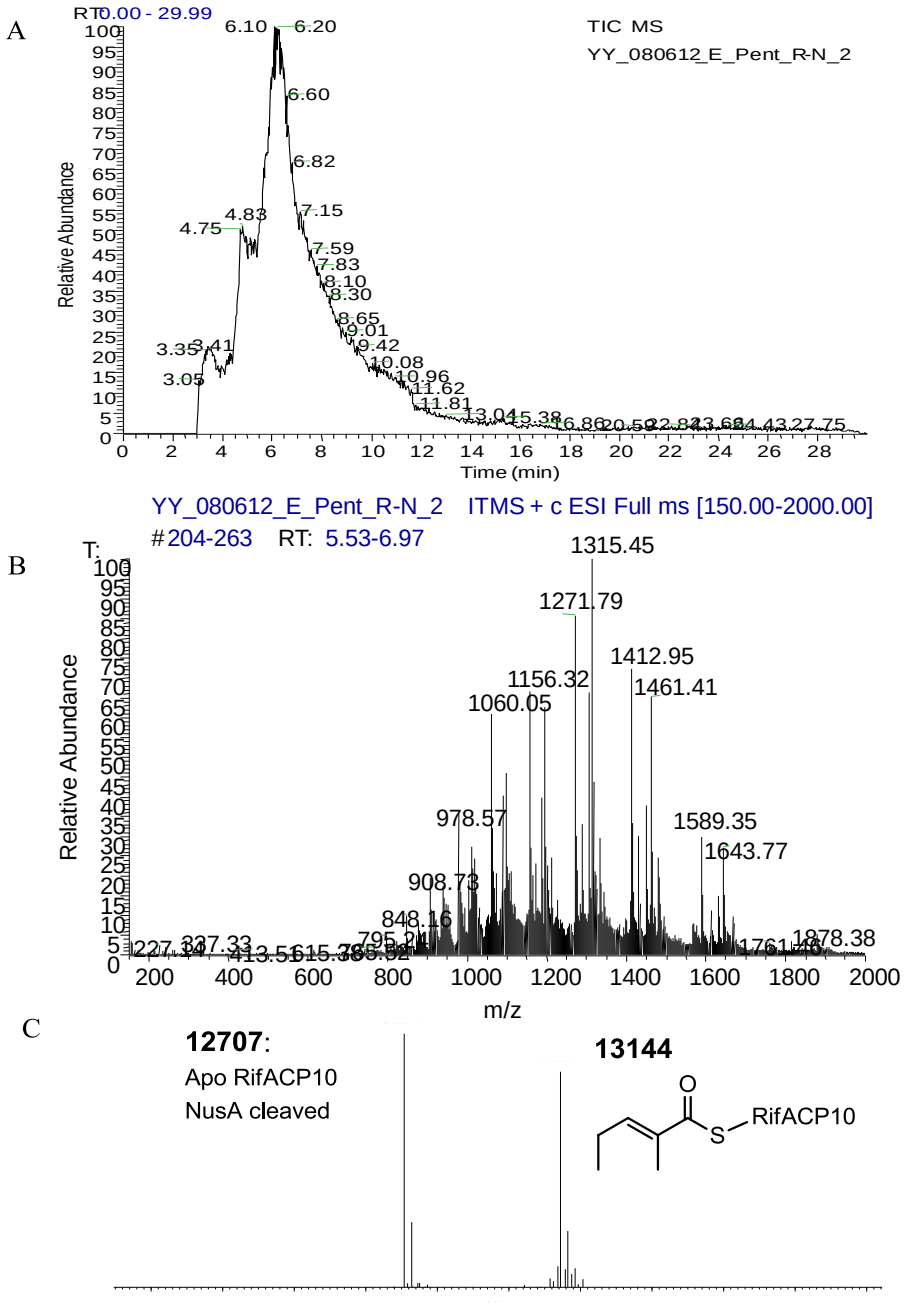


Figure 3-31. LC-ESI(+)-MS analysis of (*E*)-2-methyl-2-pentenoyl-RifACP10 generated by incubation of (*E*)-2-methylpentenoyl-CoA with Sfp and apo-RifACP10-NusA, followed by proteolytic cleavage of the NusA fragment with HRV 3C protease. **(a)** LC-MS; **(b)** LC-ESI(+)-MS; **(c)** calculated full mass of (*E*)-2-methylpentenoyl-RifACP10 (theoretical $[M+H]^+$ 13143.9 Da). *apo*-RifACP10 was also observed (theoretical mass $[M+H]^+$ 12706.7 Da).

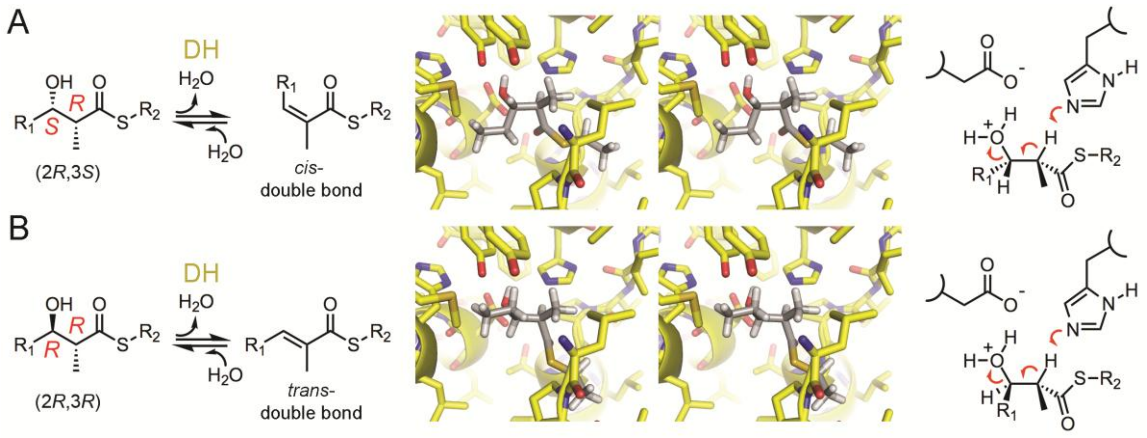


Figure 3-32. Dehydration of (2*R*,3*S*)- and (2*R*,3*R*)-2-methyl-3-hydroxyacyl thioesters. **(a)** The *syn* dehydration of a (2*R*,3*S*) intermediate would result in a *cis*-enoyl thioester. The stereodiagram shows how the C2 proton and the C3 hydroxyl group of a (2*R*,3*S*)-2-methyl-3-hydroxypentanoyl intermediate would be aligned within the active site of RifDH10 for the *syn* elimination reaction. **(b)** The *syn* dehydration of a (2*R*,3*R*) intermediate would result in a *trans*-enoyl thioester. The stereodiagram shows how the C2 proton and the C3 hydroxyl group of a (2*R*,3*R*)-2-methyl-3-hydroxypentanoyl intermediate would be aligned in the active site of RifDH10 for the *syn* elimination reaction.

Data collection	
Space group	$P3_2$
Cell dimensions, *a*, *b*, *c* (Å)	77.2, 77.2, 97.9
Resolution (Å)	50-1.82
R_{merge}	0.075 (0.807)
$I/\sigma(I)$	10.6 (2.0)
No. of reflections	55,103 (4,134)
Completeness (%)	99.1 (99.9)
Redundancy	5.3 (4.8)
Refinement	
Resolution (Å)	50-1.82
No. of reflections	55,103 (4,134)
R_{work}/R_{free}	0.211/0.254
No. of atoms	4444
Protein	4134
Water	310
Average *B* factors (Å^2)	
Protein	40
Water	46
RMS deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.759
Ramachandran Statistics (%)	
Preferred Regions	95.8
Allowed Regions	4.2
Outliers	0.0

Table 3-1. Crystallographic data and refinement statistics. Values in parentheses refer to the highest resolution shell (1.92 – 1.82 Å).

Chapter 4. Rapid modification of the pET-28 expression vector for ligation independent cloning using homologous recombination in *Saccharomyces cerevisiae*

ABSTRACT

The ability to rapidly customize an expression vector of choice is a valuable tool for any researcher involved in high-throughput molecular cloning for protein overexpression. Unfortunately, it is common practice to amend or neglect protein targets if the gene that encodes the protein of interest is incompatible with the multiple-cloning region of a preferred expression vector. To address this issue, a method was developed to quickly exchange the multiple-cloning region of the popular expression plasmid pET-28 with a ligation-independent cloning cassette, generating pGAY-28. This cassette contains dual inverted restriction sites that reduce false positive clones by generating a linearized plasmid incapable of self-annealing after a single restriction-enzyme digest. We also establish that progressively cooling the vector and insert leads to a significant increase in ligation-independent transformation efficiency, demonstrated by the incorporation of a 10.3 kb insert into the vector. The method reported to accomplish plasmid reconstruction is uniquely versatile yet simple, relying on the strategic placement of primers combined with homologous recombination of PCR products in yeast.

INTRODUCTION

Originally derived from pBR322, the pET series of cloning plasmids has been an extremely popular choice for laboratory scale inducible protein production (Studier &

Chapter 4 was adapted from the following published work: Gay G, Wagner D, Keatinge-Clay AT, and Gay DC. (2014) Rapid modification of the pET-28 expression vector for ligation-independent cloning using homologous recombination in Saccharomyces cerevisiae. *Plasmid* **76**, 66-71. Each author contributed to the experimental design and manuscript preparation.

Moffatt, 1986; Rosenberg *et al.*, 1987; Studier *et al.*, 1990). The system is currently marketed by Novagen® as the most powerful set of vectors yet developed for the subcloning and overexpression of recombinant proteins in *Escherichia coli*, with nearly 1000 scientific publications reporting the use of the pET-28 variant for protein expression in 2013†. Genes that encode a protein of interest are generally inserted into a restriction-enzyme based multiple cloning region downstream of a T7 promoter for IPTG-inducible transcription by the T7 RNA polymerase. While this gene insertion method remains a widely used technique, alternative procedures have recently gained attention for side-stepping several disadvantages inherent to restriction-enzyme based methods, including restriction-sites internal to the gene of interest and the requirement for ligase-catalyzed vector circularization.

One method for inserting genes into expression vectors that has become an increasingly popular substitute for restriction-enzyme based methods is ligation-independent cloning (LIC) (Aslanidis & Jong, 1990; Haun *et al.*, 1992). The LIC method does not require ligase to covalently circularize the vector and insert, but instead relies on the affinity of sufficiently long base-pair overhangs that anneal *in vitro* to afford a stable circular product. Base-pair overhangs of a predetermined length are generated by treatment with T4 DNA polymerase, capitalizing on the potent 3′→5′ exonuclease activity of this enzyme in the absence of free nucleotides. Primers for amplification of the gene of interest are designed to include overhangs complementary to those of the vector, avoiding complications associated with a multiple cloning region that may contain restriction sites internal to the gene. The LIC method has become especially popular

among structural biologists (Stols *et al.*, 2002; Luna-Vargas *et al.*, 2011), who often screen multiple constructs of a single protein target to determine if modifying the location of N- and C-termini impacts the crystallizability of the macromolecule. The simplicity of LIC primer design combined with the inherent amenability to parallel processing makes this cloning strategy an attractive alternative to traditional restriction-enzyme based methods for high-throughput plasmid generation.

The ability to rapidly modify an expression vector is a powerful tool that can be implemented to increase the efficiency of standard molecular cloning prevalent throughout academic and industrial laboratories. Robust methods for plasmid assembly that employ *in vivo* homologous recombination techniques have been previously reported (Ma, *et al.*, 1987, Oldenburg, *et al.*, 1997; Chino, *et al.*, 2010). We have applied these methods to construct pGAY-28, in which the multiple cloning region of the common pET-28 cloning vector was replaced with a custom LIC cassette. While Novagen® does market a limited number of pET vectors that contain LIC options for gene integration, the methods outlined here represent a rapid and uniquely versatile protocol that can be used to quickly modify any genetic element within an expression vector using homologous recombination in *S. cerevisiae*. By strategically designing primers that are positioned to anneal upstream of a unique restriction-site within the parental vector, the final step of plasmid re-circularization is greatly facilitated to enable the complete reconstruction of plasmids within a relatively short time frame (3 days). The robust nature of the new cloning vector is demonstrated through a comparison of cloning efficiencies when inserting a 10.3 kb insert into both pET-28b and pGAY-28.

RESULTS AND DISCUSSION

The pGAY-28 vector was generated to provide a LIC alternative to restriction-enzyme based cloning predominant within the pET system, while retaining all the other features found within the pET-28 subtype that has made it such a popular choice for protein overexpression in *E. coli*. The methods outlined here provide a simple and general procedure that can be adapted to reconstruct a multiple cloning region, generate a LIC cassette, or incorporate any genetic element of choice into an expression vector within several days.

The pGAY-28 expression vector was generated by amplifying the parent pET-28b(+) vector in three parts (A, B, and C), using primer pairs a_1:a_2, b_1:b_2, and c_1:c_2 (**Figure 4-1**). A key feature that permitted the rapid reconstitution of the final plasmid product involved exploiting the single XmaI site within the original pET-28 sequence. The primers a_1 and c_2 were designed to anneal upstream of the XmaI site of pET-28, so that the recombined product contained two XmaI sites adjacent to the sequences for yeast shuttle vector integration. The recombined vector (YEpADH2p / pGAY-28) becomes the DNA template for the following PCR amplification, such that primers a_1 and c_2 are used in conjunction to amplify only the pGAY-28 portion of the plasmid. The incorporation of identical restriction sites at the 5′ and 3′ ends of this amplicon facilitates the circularization of the final product that can be restricted and ligated to itself. XmaI was chosen for pGAY-28 construction to completely preserve the sequence of pET-28b(+) outside of the multiple cloning region, but any restriction site that does not exist in the parent vector can be incorporated into the primers for the same purpose. The likelihood

of selecting false positive yeast colonies is sufficiently low, and therefore the second round of PCR amplification can simply be a "colony PCR", avoiding purification and sequencing of the recombined dual vector. In this step, several yeast colonies that likely contain the recombined construct (e.g., YEpADH2p / pGAY-28) can be combined and lysed to provide template DNA for the subsequent amplification. Barring unusual recombination events, amplified DNA of the correct molecular weight in the second round of PCR is highly indicative that the recombination was successful.

The LIC cassette of pGAY-28 that replaces the multiple cloning region of pET-28 has several unique features (**Figure 4-2**). Two BseRI restriction sites within the cassette provide a single-step linearization of the vector using one restriction enzyme, generating short sticky ends that are incapable of self-annealing and susceptible to T4 DNA polymerase digestion. T4 DNA polymerase exhibits potent 3′→5′ exonuclease activity towards linear DNA, but this activity is significantly attenuated towards specific bases if high concentrations of the nucleoside triphosphate of that base exist in the reaction condition (Aslanidis & Jong, 1990). In the presence of millimolar concentrations of a specific nucleoside triphosphate, the T4 DNA polymerase will stall when it reaches that specific base in the sequence, generating considerably longer sticky ends than can be achieved with restriction enzyme digests. If BseRI linearized pGAY-28 is treated with T4 DNA polymerase in the presence of dATP, the polymerase removes nucleotides 3′→5′ from both ends of the vector until it reaches an adenine, exposing lengthy segments of single-stranded DNA. Similarly, if the amplified gene to be inserted into the LIC cassette is treated with T4 DNA polymerase in the presence of dTTP, complementary sticky ends

are generated that can anneal to the vector. Simply mixing the T4 DNA polymerase treated vector and insert before transformation is sufficient to achieve acceptable transformation efficiency of positive clones. To increase transformation efficiency further, the vector and insert should be subjected to progressive cooling, easily programmed into modern thermal cyclers. It is suspected that slowly cooling the mixture from 98 ºC to 38 ºC increases the population of complementary sticky ends that find and anneal to one another properly.

Using restriction-based methods, it has been generally observed that the size of a gene insert corresponds inversely to the number of positive clones. Our laboratory has observed a precipitous decrease in cloning efficiency with pET-28b(+) after the size of the insert exceeds 6-7 kb. To compare the cloning efficiencies of pET-28b(+) and pGAY-28, a 10.3 kb fragment (corresponding to 12 polyketide synthase domains housed within the *pksJ* gene from *B. subtilis*) was selected to challenge both systems. The gene fragment was amplified using a primer set encoding NotI and XhoI restriction sites for insertion into pET-28b(+), in addition to amplification using a primer set encoding the LIC annealing sequences for insertion into pGAY-28. Highly competent *E. coli* Top10 cells (cfu > $1 \times 10^7/\mu g$) were used to transform the constructs by electroporation, and transformants were plated on selective media. The pET-28b(+) construct yielded 10 colonies, however 5 randomly selected colonies were screened and only contained the pET-28b(+) vector without insert. The pGAY-28 construct yielded 40 colonies, and 11 of these were selected and screened. Of these, 2 were positive clones and 9 were the pGAY-28 vector without insert, indicating an approximate 18% success rate. The false positive

colonies for the pET-28b(+) construct are attributed to incomplete digestion with either NotI or XhoI, and the false positive colonies for the pGAY-28 construct are attributed to incomplete digestion with BseRI. Although 18% is a relatively low success rate, it must also be considered that the pET-28b(+) vector produced no positive clones, and that the transformation of a ~16 kb nicked vector was at least possible with pGAY-28. When smaller inserts are used to generate pGAY-28 constructs (1-2 kb), it is exceptionally rare to encounter false positive colonies due to the abundance of positive clones.

As no other features of pET-28 were modified through the incorporation of the LIC cassette, the T7 promoter, *lac* operator, ribosomal binding site, N-terminal hexahistidine tag, and thrombin cleavage site still precede the gene of interest. Translated protein products expressed from pGAY-28 will read: MGSSHHHHHHSSGLVPRGSS(#)HHHHHHH(stop), where (#) corresponds to the translation of the gene cloned into pGAY-28 for overexpression. The serine that immediately precedes (#) is the only N-terminal residue inconsistent with expression of proteins cloned into the NdeI site of pET-28b(+). While the N-terminal hexahistidine tag is mandatory and can be removed with thrombin digestion, the C-terminal heptahistidine tag is optional, and can be avoided by the incorporation of a stop codon in the reverse primer for the gene of interest.

The protocol described here can be easily adapted to modify any expression vector of choice via homologous recombination in *S. cerevisiae*. In the manner by which primers b_2 and c_1 harbor the LIC cassette used to generate pGAY-28, any sequence (and its reverse complement) can be designed into the primers to rapidly delete, amend, or add

to any region of the parent plasmid. Although the pET-28b(+) multiple cloning region that was replaced with the LIC site is relatively small, the strategy is clearly amenable to large-scale modifications, such as the exchange of drug-resistance cassettes. If longer regions (greater than 50 nucleotides) are to be incorporated, an additional set of primers will be required to first amplify the genetic material of interest. As long as the primers used for this amplification contain ends that can recombine with homologous regions of adjacent amplicons (or linearized plasmid), there is no limit to the modifications that can be achieved within a relatively short time frame. It is conceivable that the 10.3 kb insert described in this study could have been amplified with regions designed to recombine with pET-28b(+) (i.e., the annealing sequences for primers b_2 and c_1, **Figure 4-1**), transformed into yeast with the shuttle vector for recombination, followed by colony PCR of the entire 15.7 kb fragment corresponding to pET-28b(+) containing the appropriately placed insert. This fragment could subsequently be restricted with XmaI and ligated to itself to yield a construct exceptionally challenging to generate with traditional restriction-based cloning methods. Our laboratory has observed a significant increase in transformation efficiency using pGAY-28 in place of pET-28, and recently solved several crystal structures of proteins expressed from this vector (Gay *et al.*, 2014). We encourage labs engaged in high-frequency cloning coupled to protein overexpression to explore possible modifications to their vector of choice in order to increase throughput and efficiency, and recommend assembly via the protocol described herein as the most rapid and cost-effective method for this procedure.

MATERIALS AND METHODS

Strains and plasmids

E. coli TOP10 (Invitrogen) and BL21(DE3) were used as the host strains for subcloning and protein production from the pGAY-28 vector, respectively. The pET-28b(+) expression vector used as the template for plasmid modification was obtained from Novagen. *S. cerevisiae* strain BJ5464-NpgA (*MATα ura3-52 his3-Δ200 leu2-Δ1 trp1 pep4::HIS3 prb1 Δ1.6R can1 GAL*) was used for the recombinatorial assembly of PCR products (Ma *et al.*, 2009). The YEpADH2p plasmid used for colony selection in *S. cerevisiae* is a 2μ YEplac195-based shuttle vector with a *URA3* selection marker (Gietz and Akio, 1989; Ma *et al.*, 2009). Genomic DNA was extracted from *Bacillus subtilis ssp.* 168 as a template for the amplification of the 10.3 kb fragment from the *pksJ* gene.

PCR amplification of pET-28b and yeast shuttle vector integration

Unless otherwise mentioned, all DNA samples were purified by agarose gel excision using the QIAquick Gel Extraction Kit (Qiagen) and eluted into deionized water before use. Amplification of pET-28 for integration into the YEpADH2p vector was performed using standard PCR methods with KAPA HiFi polymerase (KAPA Biosystems). The vector was divided into three pieces for amplification (A, B, and C; **Figure 4-1**) using the following primers: a_1: 5′-*CAAAAAGCATACAATCAACTATCAACTATTAACTATATCGTAATACCATATG*TGTTTT CCCGGGGATCGCAG-3′, a_2: 5′-CGCATCCATACCGCCAGTTGTTTAC-3′, b_1: 5′-GAGGATGCTCACGATACGGGTTACTG-3′, b_2: 5′-

GTGTTGTGCATGA**CTCCTC**GTCTAGCGACGTGCA**GAGGAG**CTGACTGCAAA GAGCCGCGCGGCACCAG-3′, c_1: 5′-TTTGCAGTCAG**CTCCTC**TGCACGTCGCTAGAC**GAGGAG**TCATGCACAACACC ACCACCACCACCACTGAGATC-3′, and c_2: 5′-*CGCACAAATTTGTCATTTAAATTAGTGATGGTGATGGTGATGCACGTG*ATGCATGG TTACTCACCACTGCGATC-3′ (BseRI sites in bold, LIC cassette underlined, and YEpADH2p homologous recombination sites (Y-3′ and Y-5′) are italicized). YEpADH2p was restricted with NdeI and PmlI to form a linearized vector prepared for homologous recombination. The three pET-28b amplicons and linearized YEpADH2p vector were mixed to molar equivalency, and transformed into *S. cerevisiae* strain BJ5464-NpgA using the S. C. EasyComp Transformation Kit (Life Technologies). Cells were plated on uracil deficient media for selection. Many colonies (>200) appeared within 2-3 days, and six out of six colonies screened positive as successful transformants.

PCR amplification of pGAY-28 from recombined shuttle vector

Several transformed *S. cerevisiae* colonies (5-10) were transferred into 100 µl of water, and incubated at 100 ºC for 5 minutes to induce cell lysis. The cell suspension was then centrifuged at 21,000 rcf for 3 minutes, and 5 µl of the supernatant was used as the template DNA for the subsequent PCR amplification (50 µl total reaction volume), in which the primers used were a_1 and c_2. The 5′ and 3′ ends of this amplicon contained the XmaI restriction site native to the original pET-28b(+) vector, and therefore restriction with XmaI followed by ligation with T4 DNA ligase (NEB) yielded the complete pGAY-

28 expression vector that was transformed into competent *E. coli* TOP10 cells. Due to the use of a single restriction site, the number of colonies from this transformation was greater than 500. False positive clones are extremely unlikely using this method since the plasmid was generated by PCR amplification, reducing the background often observed with the transformation of doubly-digested vectors. Therefore, only two colonies were screened by DNA sequencing, and both revealed the successful construction of pGAY-28.

Standard protocol for subcloning into pGAY-28 for overexpression in E. coli

The gene of interest is amplified using the sense primer 5′-GCGGCCTGGTGCCGCGCGGCTCTAGC**(X)**-3′ and anti-sense primer 5′-GTGGTGGTGGTGGTGGTGATG**(Z)**-3′, where **(X)** represents the sequence designed to anneal to the 5′ end of the target gene, and **(Z)** represents the reversed complementary sequence designed to anneal to the 3′ end of the target gene. The LIC cassette of pGAY-28 contains dual histidine tags for purification of proteins by immobilized metal affinity chromatography, with a mandatory N-terminal hexahistidine tag and an optional C-terminal heptahistidine tag. The optional C-terminal tag can be avoided by incorporating a stop codon 5′ of the **(Z)** sequence during primer design (e.g., 5′-GTGGTGGTGGTGGTGGTGATGTTA**(Z)**-3′, stop codon underlined). Amplification of the desired gene is followed by agarose gel purification, and eluted into 50 µl of water. 2 µg of the pGAY-28 vector are restricted with BseRI (NEB), agarose gel purified, and eluted into 50 µl of water. The A_{260} for both the insert and vector are recorded.

Independently, both the linearized pGAY-28 and the target amplicon are incubated with nucleoside triphosphates in the following manner: 2 µl of 100 mM dATP is added to 50 µl of the purified vector, and 2 µl of 100 mM dTTP is added to 50 µl of the purified insert. To both (but separately), the following are then added: 11 µl water, 1 µl 0.5 M DTT, 8 µl NEB Buffer 2, 8 µl of 1 mg/ml BSA, and 2 µl T4 DNA polymerase (NEB). Both reactions are incubated at 22 ºC for 30 minutes, followed by polymerase inactivation by incubation at 75 ºC for 20 minutes. The vector and insert are then mixed in a standard 200 µl PCR tube to a total volume of 9 µl according to the following calculation:

$$\frac{V}{I} = \frac{1.9 \times A_{260_{Insert}}}{kB_{Insert} \times A_{260_{Vector}}}$$

where $\frac{V}{I}$ is the ratio of vector to insert in µl (e.g., $\frac{V}{I}$ = 2 would indicate that 6 µl of T4 DNA polymerase treated vector should be added to 3 µl of T4 DNA polymerase treated insert), $A_{260_{Insert}}$ and $A_{260_{Vector}}$ are the A_{260} values measured before T4 DNA polymerase treatment, and kB_{Insert} is equal to the kilobase pairs of the gene to be inserted into the vector. To promote the annealing of base-pair overhangs between the vector and insert, the reaction is subjected to progressive cooling using the following thermal cycler protocol: 98 ºC for 30 seconds, followed by 60 cycles (3 seconds each) where the reaction temperature is reduced by 1 ºC each cycle. While not critical for successful gene integration, this step has been observed to increase transformation efficiency of positive clones ~5 fold. After the thermal cycler reaches 38 ºC, 1 µl of 50 mM EDTA (pH 7.0) is added to the reaction mixture, and it is incubated at 22 ºC for 5

minutes. Ultimately, 1 – 5 µl of the reaction mixture can be transformed into competent *E. coli*, and plated on solid media containing kanamycin for selection of positive clones.

Comparison of cloning efficiencies between pET-28b and pGAY-28.

The *pksJ* gene from *B.* subtilis is a single ORF that encodes multiple nonribosomal peptide synthetase and polyketide synthase domains. To amplify the 10.3 kb gene fragment from *pksJ* that encodes 12 independently-folded polyketide synthase domains, primers were designed to anneal beginning with L1623 and end with the natural stop codon following V5043 (National Center for Biotechnology Information Reference Sequence: NP_389598.3). To amplify the region for insertion into pET-28b(+), the following primers were used: 5′-CATCACTTAGCGGCCGCAATGCTTGATCATATGCCGTTAACTCCGAAC-3′ and 5′-CATCACTTACTCGAGGACCTCCACCTCATAAGTATCCCATATG-3′ (NotI and XhoI restriction sites underlined). To amplify the region for insertion into pGAY-28, the following primers were used: 5′-GCGGCCTGGTGCCGCGCGGCTCTAGCCTTGATCATATGCCGTTAACTCCGAAC-3′ and 5′-GTGGTGGTGGTGGTGGTGATGTTAGACCTCCACCTCATAAGTATCCCATATG-3′ (LIC cloning regions underlined). The amplicon generated using the primers designed for incorporation into pET-28b(+) was digested with NotI and XhoI, and ligated into a pET-28b(+) vector that had been previously digested with the same restriction enzymes. The amplicon generated using the primers designed for incorporation into pGAY-28 was

treated according to the protocol described above. Both constructs were then transformed into highly competent *E. coli* cells (cfu > $1 \times 10^7/\mu g$) by electroporation.

ACKNOWLEDGEMENTS

Financial support was provided by the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin. We thank the NIH (GM106112) and the Welch Foundation (F-1712) for supporting this research. We would like to thank Dr. Yi Tang (UCLA) for sharing the BJ5464-NpgA *S. cerevisiae* strain and YEpADH2p shuttle vector.

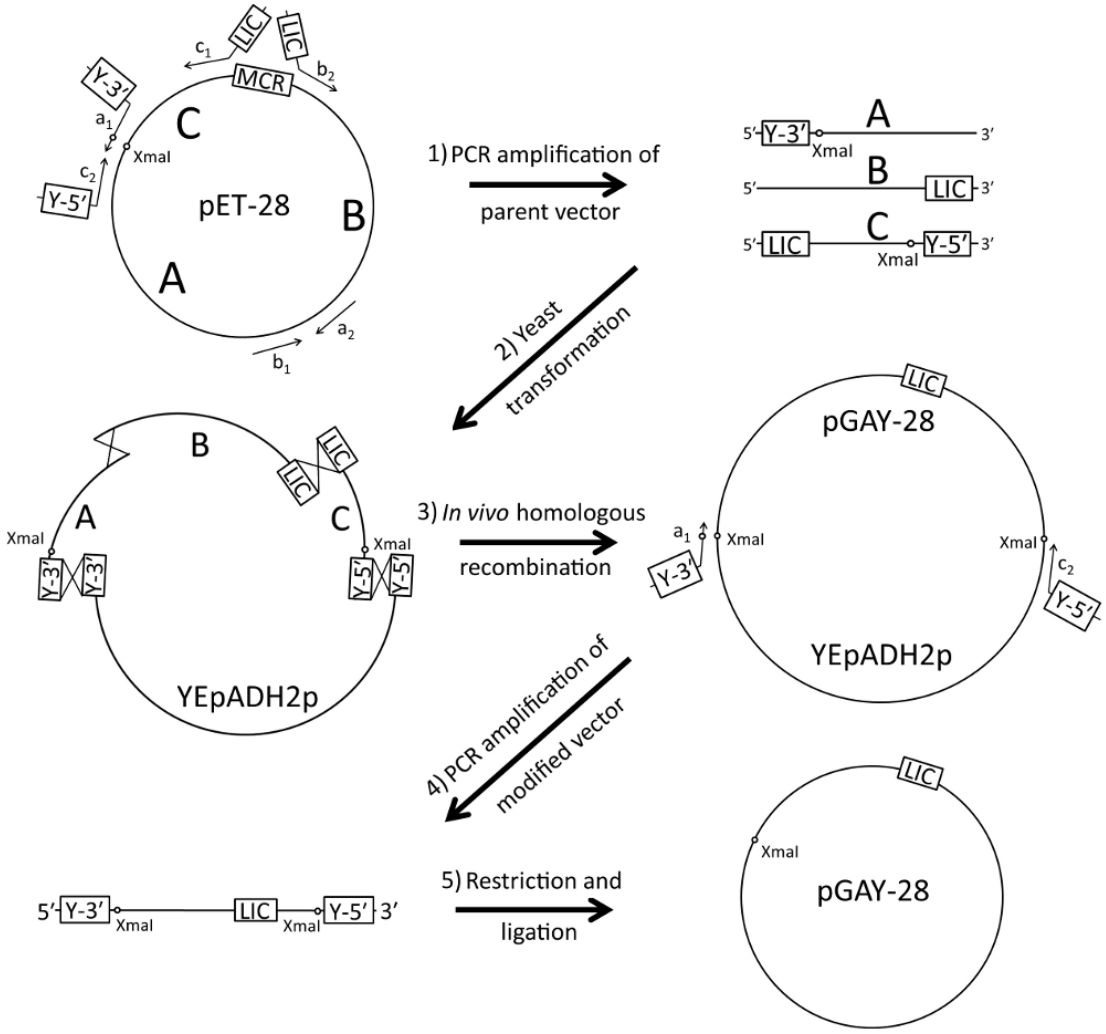


Figure 4-1. Construction of pGAY-28. The modification of pET-28 to replace the multiple cloning region (MCR) with a LIC cassette was accomplished in five steps. In step (1), the parent pET-28 vector is amplified in three segments: A, B, and C. Segment A contains a region homologous to the 3′-end of the linearized yeast shuttle vector YEpADH2p (Y-3′). Segment B contains the LIC cassette at its 3′-end. Segment C contains the LIC cassette at its 5′-end, and a region homologous to the 5′-end of YEpADH2p (Y-5′). In step (2), transformation of linearized YEpADH2p and the three amplified segments into competent *S. cerevisiae* leads to step (3), where the overlapping segments undergo homologous recombination *in vivo*. In step (4), two of the original primers from step (1) are used again to amplify the modified expression vector using "colony PCR". Since these primers were originally designed to anneal upstream of a single XmaI restriction site, step (5) involves digestion of the amplicon with XmaI followed by treatment with DNA ligase, yielding the complete pGAY-28 expression vector.

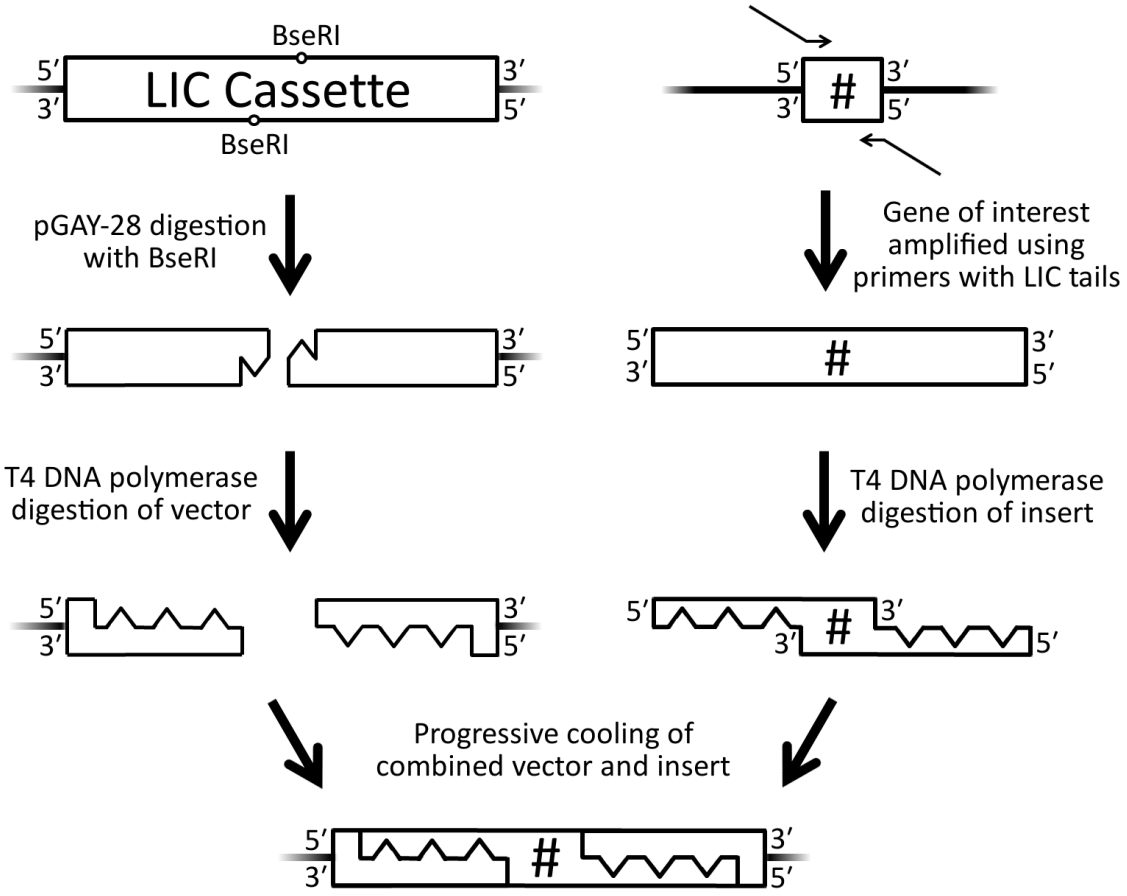


Figure 4-2. Cloning into pGAY-28 for heterologous overexpression in E. coli. To insert a target gene into pGAY-28, the vector is first digested with the restriction-enzyme BseRI. Two inverted BseRI sites ensure that the short sticky ends of the linearized vector are incapable of annealing. The gene that encodes the protein of interest (#) is PCR amplified using primers that include tails specific to the LIC integration sequence of pGAY-28. Both the linearized vector and gene insert are then digested with T4 DNA polymerase (in the presence of dATP and dTTP, respectively), generating lengthy single-stranded overhangs. The vector and insert are then combined, heated, and progressively cooled to increase the fidelity of successful gene integration (triangles on single-stranded overhangs symbolically indicate base complementarity between vector and insert).

Chapter 5. A double-hotdog with a new trick: structure and mechanism of the *trans*-acyltransferase polyketide synthase enoyl-isomerase

ABSTRACT

Many polyketide natural products exhibit invaluable medicinal properties, yet much remains to be understood regarding the machinery responsible for their biosynthesis. The recently-discovered *trans*-acyltransferase polyketide synthases employ processing enzymes that catalyze modifications unique from those of the classical *cis*-acyltransferase polyketide synthases. The enoyl-isomerase domains of these megasynthases shift double bonds and are well-represented by an enzyme that helps forge the triene system within the antibiotic produced by the prototypical bacillaene synthase. This first crystal structure of an enoyl-isomerase, at 1.73 Å-resolution, not only revealed relationships between this class of enzymes and dehydratases but also guided an investigation into the mechanism of double bond migration. The catalytic histidine, positioned differently from that of dehydratases, was demonstrated to independently shuttle a proton between the γ- and α-positions of the intermediate. This unprecedented mechanism highlights the mechanistic diversity of divergent enzymes within *trans*-acyltransferase polyketide synthases.

INTRODUCTION

Polyketides are a structurally diverse class of natural products manufactured by a broad spectrum of bacteria, fungi, and plants. The clinical importance of such polyketides

Chapter 5 was adapted from the following published work: Gay DC, Spear P, and Keatinge-Clay AT. (2014) A double-hotdog with a new trick: structure and mechanism of the *trans*-acyltransferase polyketide synthase enoyl-isomerase. *ACS Chem. Biol.* **9**, 2374-2381. Each author contributed to the experimental design and manuscript preparation.

as the antibiotic erythromycin and the cholesterol-lowering agent lovastatin has led to the structural, functional, and mechanistic investigation of the enzymatic machinery responsible for their synthesis (Keatinge-Clay, 2012). Type I polyketide synthases (PKSs) employ biosynthetic logic similar to the metazoan fatty acid synthase, which iteratively condenses and reduces acetate units into fatty acyl chains (Smith *et al.*, 2007; Maier *et al.*, 2008). In contrast to fatty acid synthesis, PKSs operate in an assembly line fashion, where modules (sets of domains responsible for a single round of condensation and processing) are structured in a linear arrangement, each obtaining a ketide extender unit selected by an acyltransferase domain (AT), condensing that extender unit to a growing polyketide chain with the ketosynthase domain (KS), and processing the resulting β-ketoacyl chain depending on which processing domains are associated with the module. These domains include a ketoreductase (KR) that catalyzes stereoselective reduction of the β-carbonyl, a dehydratase (DH) that yields an α,β-double bond through the elimination of water, and an enoylreductase (ER) that reduces the double bond to yield a completely saturated acyl intermediate.

Type I PKSs have been subdivided into two classes depending on whether ATs are encoded within modules (*cis*-AT PKSs) or discretely encoded as separate polypeptides that dock to the synthase (*trans*-AT PKSs) (Piel, 2010). While each of the domains from *cis*-AT PKSs have been structurally characterized (Keatinge-Clay, 2012), only recently has structural information become available for domains found within *trans*-AT PKS assembly lines (Bretschneider *et al.*, 2013; Gay *et al.*, 2014; Piasecki *et al.*, 2014). The earliest discovered and archetypal *trans*-AT PKS is encoded within the

Bacillus subtilis and *Bacillus amyloliquefaciens* genomes (PksX and Bae, respectively) and is responsible for the biosynthesis of the polyketide bacillaene, a polyene diamide that inhibits prokaryotic protein synthesis (Patel *et al.*, 1995). This PKS is illustrative of the unusual combinatorial logic in *trans*-AT PKS systems, which are commonly fused with non-ribosomal peptide synthetases (NRPS) and employ processing domains such as methyltransferases, enoyl-CoA hydratases, pyran synthases, and β-branching domains (B), rarely or not observed in *cis*-AT PKSs (Piel, 2010). The evolutionary acquisition of these additional processing domains has permitted *trans*-AT PKSs to explore a vast realm of possible molecular configurations, yielding natural products with functionalities not accessible to PKSs restricted to the traditional processing reactions employed in fatty acid biosynthesis.

One of the additional processing domains available to *trans*-AT PKSs is the enoyl isomerase domain (EI), which has been previously annotated as an inactive DH or DH* due to sequence alignments that predicted a DH double-hotdog fold (Kusebauch *et al.*, 2010; Moldenhauer *et al.*, 2010; Lohr *et al.*, 2013). This domain is located in what has been termed "shift-modules" that yield intermediates with β,γ-double bonds, in contrast to the classic α,β-unsaturated products of DH catalysis (**Figure 5-1**). Olefins within polyketides impact the sterically-available conformations of the final product more so than any other functional group, and therefore several studies have recently been conducted to determine if this domain is responsible for the isomerization of double bonds to positions "out-of-register" with classical PKS biosynthetic logic. Some DH domains contain intrinsic EI activity, which has been shown for DHs from the bacillaene

PKS (**Figure 5-1b**) (Moldenhauer *et al.*, 2010). Both dehydration and α,β → β,γ-double bond isomerization for C5-C6 and C7-C8 in bacillaene were demonstrated to occur within the first two DHs of PksN, without the participation of a dedicated EI domain. Whether the isomerization of the β,γ-double bond at C3-C4 of bacillaene was also catalyzed by the inherent isomerase activity of the third DH remained undetermined, and sequence alignments revealed that the following module could potentially be a shift-module.

The EI domain within the rhizoxin PKS (at the N-terminus of RhiE) has been shown to be responsible for α,β → β,γ-double bond isomerization, and a strain with an EI knockout prevented polyketide transfer downstream of the disruption (Kusebauch *et al.*, 2010). Whether this is due to KS selectivity for the properly isomerized product or interference with docking between RhiD and RhiE is unclear. A mechanism was proposed for EI-catalyzed isomerization that involved a histidine conserved between DH and EI domains, as well as an unknown residue that could act as a general base. More recently, the EI domain from the corallopyronin A synthase was isolated and studied (Lohr *et al.*, 2013). *In vitro* reconstitution of isomerase activity coupled with site-directed mutagenesis revealed that the proposed active site histidine is indeed required for catalysis; however, the unknown residue that acts in cooperation with this histidine could not be determined through mutagenic knockouts of proposed active site residues.

Here we report the 1.73-Å resolution crystal structure and reconstituted *in vitro* activity of the EI domain from the fourteenth module of the bacillaene PksX synthase (PksEI14). The structure reveals how a key active site substitution of a proline residue

conserved among PKS DH domains permits the catalytic histidine to act as both a general base and general acid in the isomerization of a double bond. Experiments conducted in deuterated solvent reveal that the mechanism proceeds through the relocation of a substrate γ-proton to the α-position without solvent interference, similar to the mechanism of double bond isomerization catalyzed by ketosteroid isomerase (Pollack, 2004). The structure also shows that EI domains may not contribute significantly to the dimerization of *trans*-AT PKS polypeptides, in contrast to what has been observed for *cis*-AT PKSs (Keatinge-Clay, 2008; Akey *et al.*, 2010; Gay *et al.*, 2013). The structural and biochemical evidence presented here reveals an unprecedented mechanism for double bond isomerization and suggests that PksEI14 is responsible for the final step in the formation of the C3-C8-triene system characteristic of bacillaene.

RESULTS AND DISCUSSION

The boundaries chosen for PksEI14 were based on the boundaries of domains upstream and downstream of PksR (PksKS14 and PksACP14, respectively). Crystals provided diffraction data enabling a model of PksEI14 to be refined to a resolution of 1.73 Å, with phasing obtained by single-wavelength anomalous dispersion from crystals soaked in methyl mercury acetate (**Table 5-1**). The structure revealed a double-hotdog fold with an overall architecture very similar to *cis*-AT PKS DH domains (C_α r.m.s.d.: 2.4 Å over 250 C_α when aligned with the DH from the fourth module of the erythromycin synthase, EryDH4; PDB Code 3EL6) (Keatinge-Clay, 2008) (**Figure 5-2a**). In contrast to

all known *cis*-AT DH structures, PksEI14 crystallized as a monomer, without the ~20-residue N-terminal region observed in those DHs that mediates their dimeric interfaces.

A structural alignment of PksEI14 with EryDH4 reveals that the α-carbon of the catalytic histidine (His18 of PksEI14) conserved in both PKS DH and EI domains superposes with the α-carbon of the catalytic histidine of EryDH4; however, the imidazole rings are offset by ~1.7 Å (as measured between N_ε atoms) (**Figure 5-2**). In PKS DHs, the catalytic histidine lies in a highly conserved HX_8P motif, in which its imidazole stacks against the proline ring (Keatinge-Clay 2008). This motif is HX_8V in PksEI14, with the $C\gamma_1$ methyl group of the valine sterically repositioning the imidazole ring. In other PKS EI domains, leucine is observed to substitute for the proline (**Figure 5-3**). The aspartic acid that completes the catalytic dyad of PKS DHs is often substituted for an asparagine in EI domains, and the structure reveals that this asparagine (Asn182) is also repositioned relative to the aspartic acid. A highly-conserved glutamine or histidine in PKS DHs located four residues upstream of the catalytic aspartic acid is presumably responsible for increasing the pK_a of this residue, priming it as a proton donor during the dehydration reaction (Gay *et al.*, 2013). The conservation of the glutamine/histidine residue is not maintained among EIs and is substituted by either a valine or leucine (Leu186 in PksEI14, **Figure 5-3**).

The previously suggested mechanism of α,β → β,γ-double bond isomerization by the EI domain involves an unknown residue that acts as a general base and abstracts a proton from the γ-carbon, similar to the isomerization mechanism for FabA (Kusebauch *et al.*, 2010; Moynié *et al.*, 2012; Lohr *et al.*, 2013). However, the structure reveals an

active site pocket dominated by hydrophobic residues that cannot participate in acid-base catalysis. The only ionizable residue within the PksEI14 active site in position to play this role is Tyr185, although it is not conserved among available EI sequences. To determine whether Tyr185 is involved in catalyzing double bond isomerization, an assay was developed to reconstitute *in vitro* activity from the isolated PksEI14 domain. As obtaining an analogue of the natural polyketide substrate to observe the reaction in the forward direction was challenging (a lengthy, highly-functionalized polyene diamide), an analogue of the reaction product, (*E*)-hex-3-enyl-*S*-pantetheine, was prepared instead. PksEI14-mediated isomerization yielding (*E*)-hex-2-enyl-*S*-pantetheine would thus resemble the anticipated reverse direction for double bond migration, namely β,γ → α,β-double bond isomerization. The reverse of the biological reaction is anticipated to be thermodynamically more favorable for this analogue since the double bond of the product is conjugated with the thioester carbonyl. Incubation of (*E*)-hex-3-enyl-*S*-pantetheine with PksEI14 followed by reversed-phase HPLC analysis revealed substrate conversion to a distinct species with a retention time equivalent to (*E*)-hex-2-enyl-*S*-pantetheine and UV absorbance maxima of 225 and 263 nm, characteristic of α,β-unsaturated thioesters (**Figure**s **5-4** and **5-5**) (Xiang *et al.,* 1999). LC/MS analysis revealed that the generated compound exhibited a molecular weight also consistent with the anticipated product. Surprisingly, the active site mutant in which the suspected tyrosine was replaced with a phenylalanine, PksEI14(Tyr185Phe), catalyzed the isomerization ~3.1-fold more rapidly, revealing that the tyrosine is not necessary for catalysis (**Figure 5-4**). The *N*-acetyl-cysteamine (NAC) derivative of the β,γ-unsaturated substrate ((*E*)-hex-3-enyl-*S*-NAC)

was also incubated with PksEI14 under similar conditions, but isomerization to (*E*)-hex-2-enyl-*S*-NAC was negligible (**Figure 5-5b**).

Since the structure does not reveal any other active site residues that could catalyze acid-base chemistry, the general base in the biological reaction could simply be a solvent-supplied water molecule, similar to the isomerization mechanism proposed for the isomerase domain of multifunctional enzyme, type-1 (Kasaragod, *et al.*, 2013). The reverse *in vitro* reaction described herein would require water to play the role of a general acid wherein the active site histidine abstracts a proton from the α-carbon of the β,γ-unsaturated substrate and a solvent water molecule donates a proton to the γ-carbon. The PksEI14-catalyzed reaction described above was conducted in D_2O to determine if a solvent-supplied deuteron would label the γ-carbon in the product, generating (*E*)-4-deutero-hex-3-enyl-*S*-pantetheine. However, in contrast to an identical reaction in H_2O, the reaction would not go to completion (as monitored by reversed-phase HPLC). After ~20-24 hours, the reaction stalled and both substrate and product slowly degraded due to the background hydrolysis of the thioester. Despite using millimolar concentrations of enzyme, insufficient isomerized product could be obtained for analysis by ^{1}H NMR. Due to the remarkably decreased rate of isomerization catalyzed by wild-type PksEI14 in D_2O, the PksEI14 mutant generated in this study (Tyr185Phe) was used for the reaction in deuterated solvent for its enhanced catalytic rate (**Figure 5-4**). It is suspected that the observed rate enhancement is the result of using a substrate analogue which does not perfectly mimic the true polyketide intermediate beyond the γ-carbon. The δ- and ε-carbons of the analogue used in this study are sp3 hybridized (unlike the natural

intermediate), and the tyrosine to phenylalanine mutation likely accommodates the bulkier substrate by increasing the volume of the active site cavity. This mutation is unexpected to alter the mechanism for isomerization from the wild-type, as it is already present in the Rhizoxin EI domain (**Figure 5-3**). The reaction catalyzed by PksEI14 (Tyr185Phe) in deuterated solvent afforded sufficient isomerized product for analysis by ^{1}H NMR (~3 mg), revealing a characteristic β-carbon proton splitting pattern that corresponded to a completely protonated methylene γ-carbon (**Figure 5-6**). As the α,β-unsaturated reaction product was not purified from the β,γ-unsaturated substrate before ^{1}H NMR analysis, (*E*)-hex-3-enyl-*S*-pantetheine was still clearly visible in the spectrum (1:0.52 molar ratio by proton signal integration). Greater than 95% of the α-protons of (*E*)-hex-3-enyl-*S*-pantetheine had exchanged for deuterons during the course of the reaction, confirming that the degree of contamination by residual solvent protons was insignificant (**Figure 5-7c**). The rate of solvent-catalyzed α-proton exchange is approximately on the same time scale as the concurrent background hydrolysis of the labile thioester in both substrate and product, complicating the analysis of reactions incubated longer than 24 hours (**Figure 5-7d**). To determine if the remarkable decrease (>10-fold) in catalytic rate of the reaction in D_2O was the result of an abnormally large kinetic isotope effect, (*E*)-2,2-dideutero-hex-3-enyl-*S*-pantetheine was incubated with PksEI14 in H_2O. Although isomerization to (*E*)-hex-2-enyl-*S*-pantetheine was observed, it could not be excluded that the product was generated primarily from a substrate species that had first undergone background exchange of α-deuterons for solvent protons, compromising any conclusions drawn from analysis of the product. It is expected that the

reduced catalytic rate observed for the reaction conducted in D_2O resulted from the exchange of deuterons for protons within PksEI14, potentially shifting the catalytic histidine out of optimal range from either the α- or γ-carbon of the substrate. It has been demonstrated that deuteration of haloalkane dehalogenase restructures the active site resulting in enzymatic inactivation (Liu, *et al.*, 2007).

To our knowledge, the only polyketide from *cis*-AT PKSs that contains a β,γ-double bond is ansamitocin (**Figure 5-1a**). Intriguingly, the ansamitocin PKS does not contain an EI domain, yet studies have shown that the formation of the β,γ-double bond occurs while the polyketide is tethered to the synthase, and the positioning of this double bond is critical for downstream transfer of the growing intermediate (Taft *et al.*, 2009). A sequence alignment of DH domains reveals that the highly conserved catalytic aspartic acid in the responsible DH (AsmDH3) is replaced by a glutamic acid, which may help enable an unusual β,ε-dehydration of the substrate. Although the migration of double bonds by polyunsaturated fatty acid synthases has also been proposed to be catalyzed by dedicated shift modules (Sossa *et al.*, 2012), no *cis*-AT PKS pathway has been discovered that employs a dedicated EI domain to catalyze α,β → β,γ-double bond isomerization.

As suggested by the results presented here, the forward biological reaction (α,β → β,γ double bond isomerization) would be thermodynamically unfavorable if the substrate did not harbor a moiety at the δ-position available for conjugation with a β,γ-double bond (e.g., double bond, amide, carbonyl). For PksEI14, this functional group is supplied by the δ,ε-double bond installed by PksDH12. This δ,ε-double bond is in conjugation with the double bond installed by PksDH11, which is in conjugation with the amide formed by

the preceding NRPS module. While the amide likely sets a thermodynamically favorable foundation for the three subsequent double bond shifts, sequence alignments do not provide a clear explanation for how PksDH11 and PksDH12 are able to catalyze both dehydration and isomerization without a dedicated EI domain. It is possible that PksDH13 is also capable of catalyzing isomerization, and that PksEI14 has been retained as a secondary checkpoint to ensure that the isomerization occurs before the polyketide is hydrolyzed from the synthase. These DHs could mediate isomerization after normal α,β-dehydration, similar to the mechanism proposed for FabA (Moynié *et al.*, 2013). Also possible is that the DHs of these modules directly dehydrate the β,γ-positions through the elimination of a γ-proton and β-hydroxyl; this type of dehydration may be facilitated by the increased acidity of the γ-proton adjacent to the amide or double bond at the δ,ε-position. In addition to the bacillaene EI domain, the EIs of both the rhizoxin and corallopyronin PKSs isomerize double bonds into conjugation with preexisting functional groups at the δ,ε position (a double bond and amide, respectively).

Our results suggest that the mechanism for double bond isomerization by the EI domain proceeds via the shuttling of a substrate proton by a single active site histidine, similar to the mechanism of triose phosphate isomerase (Plaut and Knowles, 1972). An alignment of the six crystal structures of *cis*-AT PKS DHs solved to date reveal that the imidazole groups of the catalytic histidines superpose nearly perfectly (Gay *et al.*, 2013), yet the imidazole group of PksEI14 is relatively shifted ~1.7 Å (as measured between N_ε atoms) (**Figure 5-2b**). Since the EI domain possesses the same double-hotdog fold as the DH domain, the ACP-bound substrate may be presented to the active site residues of

these two domains in a similar manner. The shifting of the catalytic histidine from its position in DH domains is approximately the distance of a carbon-carbon bond and could position the imidazole group between the sites of proton abstraction and donation. The shuttling of a substrate proton catalyzed by a single residue to isomerize a double bond has been proposed for several other isomerases, including ketosteroid isomerase (Pollack, 2004), human mitochondrial enoyl-CoA isomerase (Partanen *et al.*, 2004), cyclohexenylcarbonyl-CoA isomerase (Reynolds *et al.*, 1993), and $\Delta^{3,5}$,$\Delta^{2,4}$-dienoyl-CoA isomerase (Zhang *et al.*, 2001).

The possible intermediate of the mechanism proposed here involves either the transient protonation of the thioester carbonyl to an enol or the formation of an enolate that is stabilized by an oxyanion hole. To better understand how the single active site histidine catalyzes isomerization unaided and what the intermediate may be, substrate and product analogues were modeled into the active site of PksEI14. The location of the substrate and product were subjected to several restraints. First, a conserved positively-charged residue and phenylalanine have been suggested to form the ACP-docking site for *cis*-AT PKS DHs (Keatinge-Clay, 2008), and these residues are retained in PksEI14 (Lys257 and Phe204). The modeled substrate and product were therefore oriented so that the phosphopantetheinyl arm would be in close proximity to these residues. Second, since the active site histidine must interact with both the α- and γ-carbons of the polyketide, a second restraint limited the distance of these carbons to within 3.0 angstroms of the ε-nitrogen of the catalytic histidine. Finally, the remaining atoms were manually adjusted to minimize steric clash. These restraints revealed that the thioester carbonyl would be

positioned to replace Wat2, forming hydrogen bonds with both the backbone amide of Gly28 and Wat3 (**Figure 5-8**). Gly28 forms the N-terminus of the first central α-helix, which contributes a partial positive charge from the helical dipole. Wat3 is coordinated by the carboxamide side chain of Asn182 (which replaces the catalytic aspartic acid of PKS DHs) and the backbone amide of Ala29, the second residue of this central helix (**Figure 5-2c**). An oxyanion hole formed by these two interactions could stabilize the thioester in an enolate form. No ionizable side chain is available in this region to donate a proton to the thioester carbonyl to form an enol intermediate. The participation of water molecules as oxyanion hole hydrogen bond donors is quite common amongst enolizing enzymes (Pihko *et al.*, 2009). If the model presented in **Figure 5-8** is an accurate representation of how the substrate binds to the active site of PksEI14, the stereoselectivity of the *in vivo* reaction would generate an *S* stereochemistry at the methyl-bearing α-carbon (C2 of bacillaene).

In contrast to each of the structures of *cis*-AT PKS DHs that have been solved to date, PksEI14 crystallized as a monomer. This raises questions regarding the architecture of *trans*-AT PKS shift-modules, represented by module 14 of PksX (generally composed of a condensation-incompetent KS, an EI, and an ACP domain). *cis*-AT PKS DHs dimerize primarily through interactions between their 20 N-terminal residues (Keatinge-Clay 2008; Akey *et al.*, 2010; 2013; Gay *et al.*, 2013). However, an equivalent N-terminal stretch is absent in both PksEI14 and the homologous domain from the Bae synthase. The recently published structure of the KS-B didomain from the rhizoxin *trans*-AT PKS reveals that the B domain, structurally homologous to a DH domain, dimerizes

quite differently from *cis*-AT PKS DHs with a less extensive interface (Bretschneider *et al.*, 2013). This smaller interface may also be the only interaction between the EI domains of shift modules.

This study reports the first structure of the *trans*-AT PKS EI domain, revealing how divergent evolution of the double-hotdog fold has rendered an additional PKS domain capable of distinct chemistry (Pidugu *et al.*, 2009; Lohman *et al.*, 2013, He *et al.*, 2014). The results presented here strongly suggest that a key active site modification between PKS DH and EI domains enables the EI histidine to shuttle a proton between the α- and γ-carbons of an enoyl intermediate. Characterization of the reconstituted i*n vitro* activity catalyzed by PksEI14 in deuterated solvent supports a mechanism whereby a γ-proton (C4 of bacillaene) is abstracted by the active site histidine and is relocated to the α-position (C2) without the participation of solvent protons. To our knowledge, this enzymatic mechanism is the first example of a histidine independently shuttling a substrate proton to catalyze α,β → β,γ-double bond isomerization.

MATERIALS AND METHODS

Cloning, expression, and purification of PksEI14

Domain boundaries for the polypeptide fragment containing PksEI14 were determined based on sequence alignments of the previous and subsequent domains (PksKS14 and PksACP14, respectively). The N-terminus was chosen to be Ser1124 of PksR, 10 residues after the last conserved residue of PksKS14 (Phe1113), and the C-

terminus was chosen to be Asn1395, 17 residues before the first conserved residue of PksACP14 (Ile1413). The DNA encoding PksEI14 was amplified using primers 5'-GCGGCCTGGTGCCGCGCGGCTCTAGCGATGTTGAAGACAATATCAATCAG-3' and 5'-GTGGTGGTGGTGGTGGTGATGTTAATTCGAAATACGAAGCTGTTTAAG-3' (ligation independent cloning regions underlined) from *Bacillus subtilis subsp*. 168 and inserted into pGAY28b, a ligation-independent cloning vector constructed from pET28b (Gay *et al.*, 2014). *E. coli* BL21(DE3) transformed with the expression plasmid was inoculated into LB media containing 50 mg/L kanamycin at 37 °C, grown to OD_{600} = 0.4, and induced with 0.5 mM IPTG. After 12 h at 15 °C, cells were collected by centrifugation and resuspended in lysis buffer (0.5 M NaCl, 10% (v/v) glycerol, 0.1 M HEPES, pH 7.5). Following sonication, cell debris was removed by centrifugation (30,000 x g, 30 min). The supernatant was poured over a column of Nickel-NTA resin (Thermoscientific), which was then washed with 50 mL lysis buffer containing 15 mM imidazole and eluted with 5 mL lysis buffer containing 150 mM imidazole. PksEI14 was further purified using a Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated with 150 mM NaCl, 10 mM HEPES, pH 7.5. The eluted protein was concentrated to 30 mg/mL in the equilibration buffer and stored at -80 ˚C until needed.

The PksEI14(Tyr185Phe) expression plasmid was generated by PCR amplification from the PksEI14 construct described above using primers 5'-ATGCGACATGAATTCTGCGTTCTTGGCTATCCTCAGTTTCTTAGAGCAGTTTG ACATGAC-3' and 5'-

TCATGATCGGAATTCGTCATCAATGGGCTTACAGTGTAATGGTGCCCGTTTGC CAGCCCGC -3’ (EcoRI site underlined). The PCR product was digested with EcoRI and ligated. Purification of the mutant protein was performed as described for unmutated PksEI14.

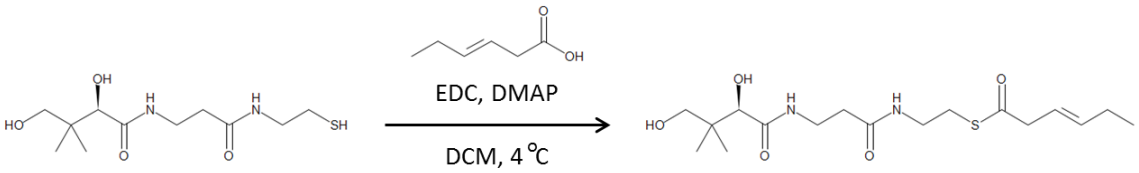
Cloning, expression, and purification of 4-coumaroyl-CoA-ligase (CCL)

The DNA encoding CCL was amplified using primers 5’-TCACTTAGACATATGTTCCGCAGCGAGTACGCA-3’ and 5’-TGGACGTTAACTCGAGTCATCGCGGCTCCCTGAGCTGT-3’ (NdeI and XhoI restriction sites underlined) from *S. coelicolor* A3(2) and inserted into pET28b. *E. coli* BL21(DE3) transformed with the expression plasmid was inoculated into LB media containing 50 mg/L kanamycin at 37 °C, grown to OD_{600} = 0.4, and induced with 0.5 mM IPTG. After 12 h at 15 °C, cells were collected by centrifugation and resuspended in lysis buffer (0.5 M NaCl, 10% (v/v) glycerol, 0.1 M HEPES, pH 7.5). Following sonication, cell debris was removed by centrifugation (30,000 x g, 30 min). The supernatant was poured over a column of Nickel-NTA resin (Thermoscientific), which was then washed with 50 mL lysis buffer containing 15 mM imidazole and eluted with 5 mL lysis buffer containing 150 mM imidazole. CCL was further purified using a Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated with 150 mM NaCl, 10 mM HEPES, pH 7.5. The eluted protein was concentrated to 30 mg/mL in the equilibration buffer and stored at -80 ˚C until needed.

Crystallization, structure determination, and refinement

Crystals of PksEI14 grew in 1-4 d by sitting drop vapor diffusion at 22 °C. Drops were formed by mixing 3 μL of protein solution (30 mg/mL PksEI14, 150 mM NaCl, 10 mM HEPES, pH 7.5) with 1 μL of crystallization buffer (1.9 M ammonium sulfate, 0.1 M Tris-HCl, pH 8.3). Crystals were frozen in liquid nitrogen after a 5-min soak in the crystallization buffer modified with 10% (v/v) glycerol. Diffraction data, collected at APS GMCA Beamline 23-ID-D, were processed by HKL2000 (Otwinowski & Minor, 1997). The structure was initially solved to 1.88-Å resolution by single-wavelength anomalous dispersion (**Table 5-1**). To generate a heavy atom derivative data set for phasing, crystals of PksEI14 were grown as described above and frozen in liquid nitrogen after a 30-min soak in the crystallization buffer modified with 10% (v/v) glycerol and 3 mM methyl mercuric acetate. Diffraction data for derivative crystals were collected at ALS Beamline 5.0.2, and processed by HKL2000. Heavy atom sites were located and an initial model generated using AutoSol and AutoBuild within the Phenix suite (Terwilliger *et al.*, 2008; Terwilliger *et al.*, 2009). The initial model was then refined using a data set collected on a native crystal to 1.73 Å resolution, by iteratively building into the remaining electron density map with Coot (Emsley *et al.*, 2010) and refining with Refmac5 (Murshudov *et al.*, 1997).

Synthesis of thioester substrates



(E)-hex-3-enyl-S-pantetheine: A solution of D-pantetheine (90 mg, 0.35 mmol, 1 eq.) dissolved in 5 ml DCM was set to stir at 4 ºC. To this solution was added (*E*)-hex-3-enoic acid (40 mg, 0.35 mmol, 1eq.), EDC (67.1 mg, 0.35 mmol, 1 eq.), and DMAP (4 mg, 0.035 mmol, 0.1 eq.) and was set to stir at 4 ºC for 1 h and then at 22 ºC for an additional hour. Solvent was removed *in vacuo*, and the remaining material was chromatographed on silica gel eluting with 1:1 EtOAc:acetone to afford (*E*)-hex-3-enyl-*S*-pantetheine (100 mg, 81%) as a clear oil.

R_f = 0.3 (1:1 EtOAc:acetone)

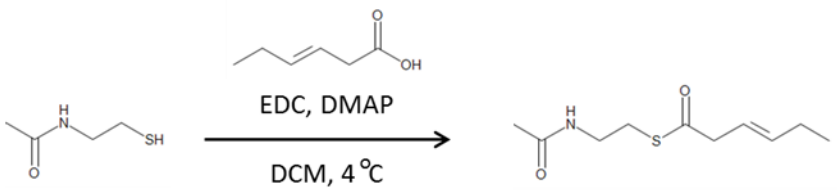
^{1}H NMR: 400 MHz ($CDCl_3$) δ = 7.34 (t, *J* = 5.3 Hz, 1H), 6.16 (t, *J* = 5.3 Hz, 1H), 5.71-5.63 (m, 1H), 5.51-5.43 (m, 1H), 3.99 (d, *J* = 5.2 Hz, 1H), 3.82 (d, *J* = 5.2 Hz, 1H), 3.59-3.32 (m, 6H), 3.24 (dq, J_1 = 7.0 Hz, J_2 = 1.1 Hz, 2H), 3.08-2.95 (m, 2H), 2.41 (t, *J* = 5.8 Hz, 2H), 2.11-2.03 (m, 2H), 1.02 (s, 3H), 0.99 (t, *J* = 7.4 Hz, 3H), 0.92 (s, 3H).

LC/MS: Calc. for $C_{17}H_{31}N_2O_5S$ $[M+H]^+$: 375.5. Found 357.2 $[M+H-H_2O]^+$, 375.2 $[M+H]^+$, and 397.2 $[M+Na]^+$.

(E)-hex-2-enyl-S-pantetheine: Prepared identically to (*E*)-hex-3-enyl-*S*-pantetheine, substituting (*E*)-hex-2-enoic acid for (*E*)-hex-3-enoic acid.

R_f = 0.3 (1:1 EtOAc:acetone)

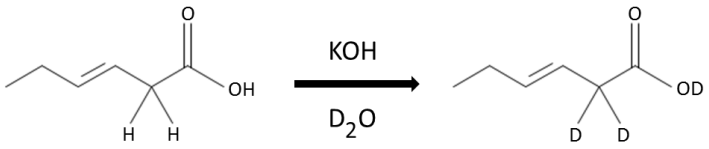
^{1}H NMR: 600 MHz ($CDCl_3$) δ = 6.90 (dt, J_1 = 15.5 Hz, J_2 = 7.4 Hz, 1H), 6.10 (dt, J_1 = 15.5 Hz, J_2 = 1.6 Hz, 1H), 3.56-3.34 (m, 7H), 3.13-3.01 (m, 2H), 2.39 (t, J = 5.9 Hz, 2H), 2.17 (qd, J_1 = 7.6 Hz, J_2 = 1.5 Hz, 2H), 1.48 (m, 2H), 1.00 (s, 3H), 0.92 (t, J = 7.4 Hz, 3H), 0.90 (s, 3H).



(E)-hex-3-enyl-S-NAC: This compound was prepared identically to *(E)-hex-3-enyl-S-pantetheine*, replacing 90 mg of D-pantetheine with 42 mg of *N*-acetyl-cysteamine (NAC) (0.35 mmol, 1 eq.). The crude material was chromatographed on silica gel eluting with EtOAc to afford (*E*)-hex-3-enyl-*S*-NAC (55 mg, 67%) as a white powder. The compound was dissolved in 1.3 ml of 3:1 DMSO:H_2O to provide a 200 mM stock solution.

R_f = 0.5 (EtOAc)

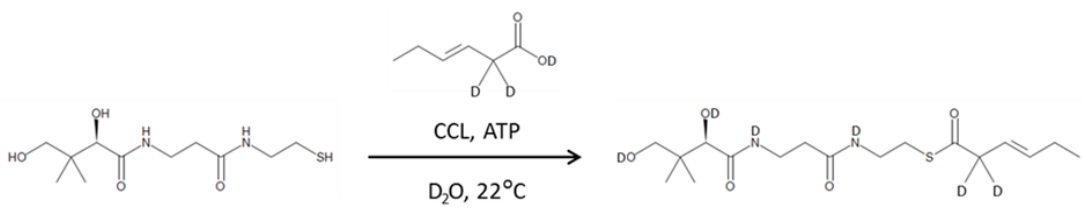
^{1}H NMR: 400 MHz ($CDCl_3$) δ = 5.67 (m, 1H), 5.49 (m, 1H), 3.43 (q, *J* = 6.3 Hz, 2H), 3.24 (d, *J* = 7.0 Hz, 2H), 3.01 (t, *J* = 6.5 Hz, 2H), 2.07 (t, *J* = 5.6 Hz, 2H), 1.96 (s, 3H), 1.00 (t, J_1 = 7.4 Hz, 3H).



(E)-2,2-dideutero-hex-3-enoic acid: To a 5 ml solution of 5 M KOH dissolved in D_2O, (*E*)-hex-3-enoic acid was added (200 mg). The solution was set to stir at 50 ºC for 1 h, after which the solvent was removed *in vacuo*. An additional 5 ml of D_2O was added to solution, which was again set to stir at 50 ºC for 1 h, and the solvent was removed *in vacuo*. The process of adding fresh D_2O followed by removal *in vacuo* was repeated a total of four times. The solution was then acidified with 10 M HCl to a pH of approximately 2, and extracted three times with EtOAc. Solvent was removed *in vacuo*, to afford (*E*)-2,2-dideutero-hex-3-enoic acid (195 mg, 98%).

R_f = 0.8 (75:24:1 hexanes:EtOAc:acetic acid)

^{1}H NMR: 400 MHz ($CDCl_3$) δ = 5.64 (dt, J_1 = 15.3 Hz, J_2 = 6.3 Hz, 1H), 5.49 (d, *J* = 15.3 Hz, 1H), 2.11-2.02 (m, 2H), 0.99 (t, *J* = 7.4 Hz, 3H).



(E)-2,2-dideutero-hex-3-enoyl-S-pantetheine: To a round bottom flask, the following were added: 5 ml of 1 M HEPES (dissolved in H_2O), 284 mg of $MgCl_2 \bullet 6H_2O$, 1 ml of glycerol, and 1 ml of 0.35 M D-pantetheine (dissolved in H_2O, 0.35 mmol, 1 eq.), pH 7.5. Solvent was removed *in vacuo*, and the contents were resuspended in 5 ml D_2O. This was set to stir at 50 ºC for 1 h, after which the solvent was removed again *in vacuo*. The process of removing the solvent *in vacuo* and resuspending in D_2O was repeated two additional times. The contents were then resuspended in 15 ml D_2O, and the following were added: 300 mg NaCl, 355 mg adenosine triphosphate disodium salt, (*E*)-2,2-dideutero-hex-3-enoic acid (50 mg, 0.44 mmol, 1.3 eq.), and 200 µl of purified CCL (30 mg/ml, 150 mM NaCl, 15 mM HEPES, pH 7.5, H_2O). This reaction was allowed to incubate 16 h at 22 ºC, and the following day was extracted twice with 20 ml of EtOAc, which was removed *in vacuo* to afford (*E*)-2,2-dideutero-hex-3-enyl-*S*-pantetheine (85 mg, 65%).

R_f = 0.3 (1:1 EtOAc:acetone)

^{1}H NMR: 400 MHz ($CDCl_3$) δ = 7.34 (t, *J* = 5.3 Hz, 1H), 6.16 (t, *J* = 5.3 Hz, 1H), 5.64 (dt, J_1 = 15.3 Hz, J_2 = 6.3 Hz, 1H), 5.50 (d, *J* = 15.3 Hz, 1H), 3.99 (d, *J* = 5.2 Hz, 1H),

3.82 (d, *J* = 5.2 Hz, 1H), 3.59-3.32 (m, 6H), 3.08-2.95 (m, 2H), 2.41 (t, *J* = 5.8 Hz, 2H), 2.11-2.03 (m, 2H), 1.02 (s, 3H), 0.99 (t, *J* = 7.4 Hz, 3H), 0.92 (s, 3H).

LC/MS: Calc. for $C_{17}H_{29}D_2N_2O_5S$ $[M+H]^+$: 377.5. Found 359.2 $[M+H-H_2O]^+$, 377.2 $[M+H]^+$, and 399.2 $[M+Na]^+$.

Enzymatic activity assays

***In vitro* reconstitution of PksEI14 activity in H_2O.**

A master mix was first generated, composed of the following: 8 µl from a stock of 100 mM (*E*)-hex-3-enyl-*S*-pantetheine, 20 µl of glycerol, 20 µl of 1 M HEPES (pH 7.5), 20 µl of 4 M NaCl, and 132 µl of water. This was split into three 40 µl reactions, to which were added either (1) 40 µl of protein equilibration buffer (150 mM NaCl, 10 mM HEPES, pH 7.5), (2) 40 µl of purified PksEI14, or (3) 40 µl of purified PksEI14(Tyr185Phe). The final concentration of each component would be: 2 mM (*E*)-hex-3-enyl-*S*-pantetheine, 5% glycerol (v/v), 55 mM HEPES (pH 7.5), 275 mM NaCl, and 15 mg/ml (460 µM) of PksEI14 for reaction (2) and 15 mg/ml (460 µM) of PksEI14(Tyr185Phe) for reaction (3). The reactions were incubated for 24 h at 37 ºC, after which they were heat-treated at 80 ºC for 5 min and centrifuged at 21,000 x g for 10 min to deactivate and precipitate the enzyme. The reactions were then injected onto a C_{18} reversed-phase HPLC column and monitored over a range of λ values ranging from 200-600 nm (7:3 water:MeOH, 0.1 % TFA, over 45 min). A peak with a λ_{max} of 225 and 263 nm (characteristic of α,β-unsaturated thioesters) that was not observed in control reaction (1) was observed in both reactions (2) and (3), which contained wild-type and mutant

PksEI14, respectively. The new peak was collected and subjected to ^{1}H NMR and LC/MS, confirming its identity as (*E*)-hex-2-enyl-*S*-pantetheine (Calc. for $C_{17}H_{31}N_2O_5S$ $[M+H]^+$: 375.5. Found 357.2 $[M+H-H_2O]^+$, 375.2 $[M+H]^+$, and 397.2 $[M+Na]^+$). Reconstitution of enzymatic activity towards (*E*)-hex-3-enyl-*S*-NAC was carried out in an identical manner to the reaction using (*E*)-hex-3-enyl-*S*-pantetheine, only replacing the 2 mM (*E*)-hex-3-enyl-*S*-pantetheine with 2 mM (*E*)-hex-3-enyl-*S*-NAC. The 200 mM stock solution of (*E*)-hex-3-enyl-*S*-NAC was dissolved in 3:1 DMSO:H_2O, and therefore the final concentration of each component would be: 2 mM (*E*)-hex-3-enyl-*S*-NAC, 5% glycerol (v/v), 55 mM HEPES (pH 7.5), 275 mM NaCl, 0.75% (v/v) DMSO, and 15 mg/ml (460 μm) PksEI14.

PksEI14 isomerase assay conducted in deuterated solvent.

To exchange PksEI14 into deuterated solvent, a deuterated isomerase buffer (DIB) was first prepared. 10 ml of glycerol, 1.2 g NaCl, and 0.5 g MES were dissolved into 10 ml of 99.9% D_2O, and the pH/pD set to 7.0 with 5 M NaOH. This mixture was then incubated at 60 ˚C for 30 minutes, and dried under vacuum until the mass of the container holding the contents no longer decreased after additional drying. An additional 10 ml of D_2O was added, and the mixture was again subjected to incubation at 60 ˚C for 30 minutes followed by drying under vacuum. This process was repeated an additional 3 times, after which the dried contents were resuspended in 90 ml D_2O to generate DIB (10% glycerol (v/v), 200 mM NaCl, 25 mM MES, pD 7.0, >98% D_2O).

Due to the remarkably decreased rate of isomerization catalyzed by PksEI14 in D_2O, the PksEI14 mutant generated in this study (Tyr185Phe) was chosen for the reaction in deuterated solvent for its enhanced catalytic rate (**Figure 5-4**). This mutation is highly unlikely to alter the mechanism for isomerization from the wild-type, as it is already present in the Rhizoxin EI domain (**Figure 5-3**). To exchange PksEI14 (Tyr185Phe) into DIB, the hexa-histidine tagged protein was first passed over a column of Ni-NTA agarose resin. The bound protein was washed with 10 ml of lysis buffer, and then washed with 10 ml of DIB. The protein was then eluted from the resin with 5 ml of DIB modified by the addition of pure imidazole (150 mM final concentration). To reduce the concentration of both imidazole and any remaining solvent protons, the 5 ml of eluted protein solution was diluted with 5 ml of DIB to a total volume of 10 ml, followed by concentration to a volume of 5 ml using a centrifugal protein concentrator with a 30 kDa MWCO filter (Thermo-Scientific). The process of diluting the protein solution with DIB followed by concentration was repeated 10 times, effectively reducing the imidazole concentration to the low micromolar range. The protein was then concentrated to 2 mM.

To initiate the title assay, 10 mg of pure (*E*)-hex-3-enyl-*S*-pantetheine was dissolved into 2 ml of DIB, to which was added 2 ml of the deuterated PksEI14 (Tyr185Phe) solution described above, giving a final concentration of each reaction component as: 10% glycerol (v/v), 200 mM NaCl, 25 mM MES, 7 mM (*E*)-hex-3-enyl-*S*-pantetheine, 1 mM PksEI14 (Tyr185Phe), <1 mM imidazole, >98% D_2O, pD 7.0. As a negative control, 10 mg of pure (*E*)-hex-3-enyl-*S*-pantetheine was simultaneously dissolved into 4 ml of DIB. The reaction was set to incubate at 22 ºC for up to 48 h,

monitored by HPLC in a similar manner to the activity assay performed in H_2O. It was noted that the isomerization activity progressively slowed during the time course of the reaction, presumably from the background exchange of the substrate α-protons for deuterons. After 20-24 h, the rate of isomerization became slower than the rate of thioester hydrolysis, leading to the simultaneous stalling of the enzymatic reaction and degradation of both substrate and product. The reaction was therefore quenched after 24 h by extraction with EtOAc (3 x 10 ml). The extracts were then dried under vacuum and resuspended in 1 ml $CDCl_3$ for ^1H NMR analysis, revealing two points of interest. First, the α-proton signal (3.24 ppm) for the unreacted β,γ-unsaturated substrate was absent in the spectra and the shifts associated with the substrate olefin protons (5.71-5.63 and 5.51-5.43) revealed identical splitting patterns as observed for (*E*)-2,2-dideutero-hex-3-enyl-*S*-pantetheine (**Figure**s **5-6** and **5-7**). Second, the splitting pattern for the β-proton signal of the α,β-unsaturated isomerized product was a doublet of triplets, indicative of a protonated γ-carbon. Residual contamination of the reaction with solvent protons is an unlikely source for the γ-protons in the isomerized product, since the background conversion of (*E*)-hex-3-enyl-*S*-pantetheine to (*E*)-2,2-dideutero-hex-3-enyl-*S*-pantetheine confirms that the reaction solvent was greater than 98% D_2O.

PksEI14 catalyzed isomerization of (*E*)-2,2-dideutero-hex-3-enyl-*S*-pantetheine.

A reaction identical to the assay described above where the isomerase activity of PksEI14 was reconstituted in H_2O was initiated; however, an α-deuterated version of the substrate ((*E*)-2,2-dideutero-hex-3-enyl-*S*-pantetheine) was substituted for (*E*)-hex-3-

enyl-*S*-pantetheine. The reactions were monitored by HPLC as described above, but a remarkable decrease in the rate of PksEI14-catalyzed isomerization compounded with the background exchange of α-substituted protons (or deuterons) of the β,γ-unsaturated thioester with solvent complicated analysis of reaction products (i.e., the small amount of isomerized product which did form could have arisen either by 1) the direct enzymatic shuttling of an α-deuteron to the γ-position, or 2) the initial background exchange of the α-deuteron for a proton, followed by enzyme catalyzed shuttling of that proton to the γ-position) (**Figure 5-7d**).

^{1}H NMR analysis was performed on a Varian Mercury 400 MHz or a Varian DirectDrive 600 MHz instrument. LC-MS was performed on an Agilent Technologies 1200 Series HPLC with a Gemini C_{18} column (5 μm, 2 × 50 mm; Phenomenex) coupled to an Agilent Technologies 6130 quadrupole mass spectrometer system equipped with an electrospray-ionization source. A 5–95% B gradient over 12 min at a flow rate of 0.7 ml/min was run in which the mobile phases were (A) H_2O with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid. Reversed-phase HPLC was performed using an isocratic flow rate of 1 ml/min through a C_{18}-column, with 34.95% MeOH/64.95% H_2O/0.1% TFA as the mobile phase, monitoring UV-absorbance at wavelengths ranging from 200-600 nm.

ACKNOWLEDGEMENTS

Funding was provided by the National Institutes of Health (NIH; GM106112). Instrumentation and technical assistance for crystallographic work were provided by A.

Monzingo and the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin. The Berkeley Center for Structural Biology is supported in part by the NIH, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. We would like to thank A. Hughes for generating the CCL expression plasmid.

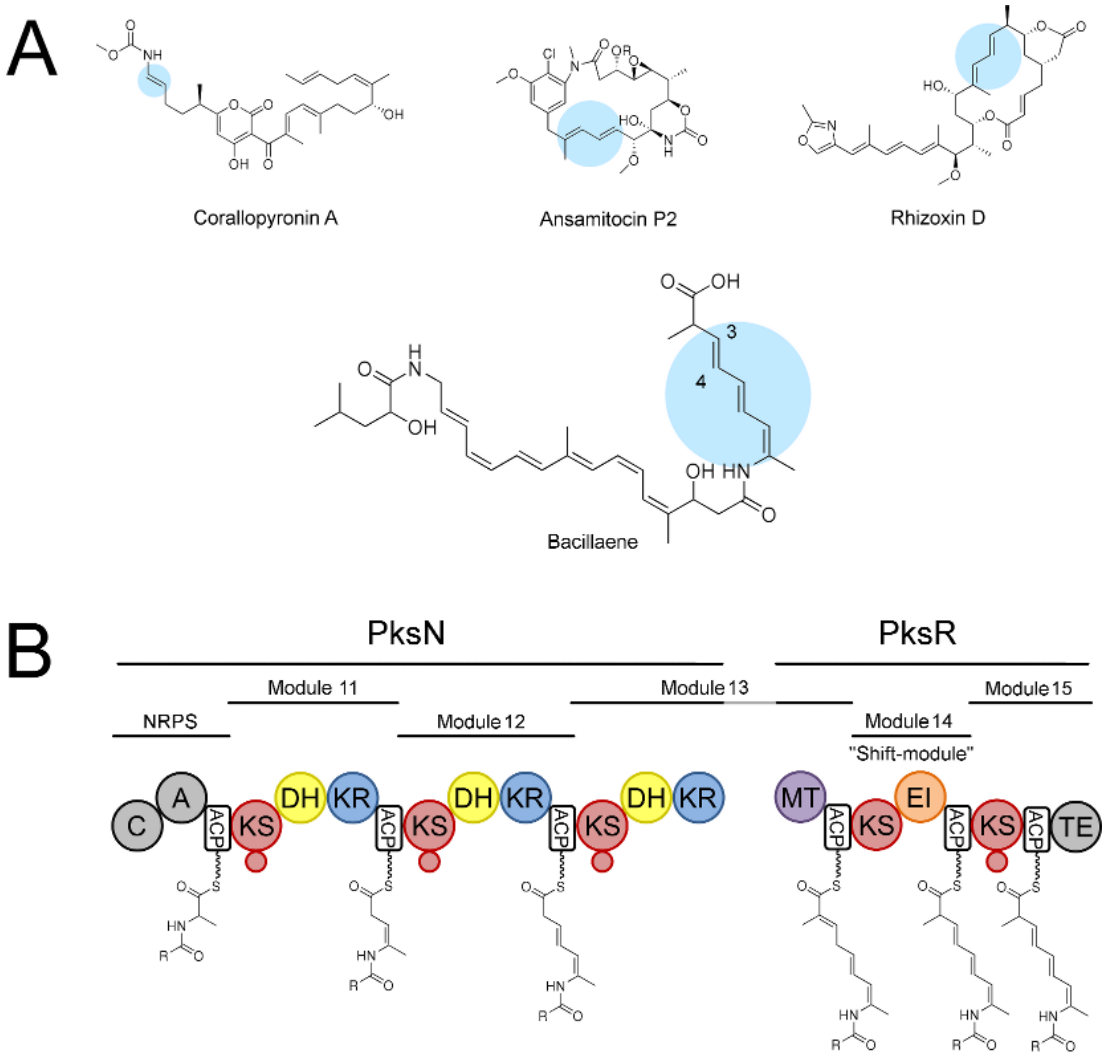


Figure 5-1. Polyketides with shifted double bonds. **(a)** Several polyketides with isomerized double bonds are highlighted with blue circles. The C3/C4 double bond of bacillaene, isomerized by PksEI14, is annotated. **(b)** The last two PKS subunits responsible for the final catalytic steps of bacillaene biosynthesis are PksN and PksR. The proposed route involves the installation of β,γ-double bonds by the DH domains of modules 11 and 12 (PksDH11 and PksDH12), while the EI domain of module 14 is a dedicated isomerase for the α,β-double bond installed by the DH from module 13. The KS domain of the terminal module (PksKS15) is not condensationally competent and likely acts as a gatekeeper to ensure that the isomerization has taken place before the final product is released from the synthase. The small red circles appended to the KS domains indicate the presence of a flanking subdomain.

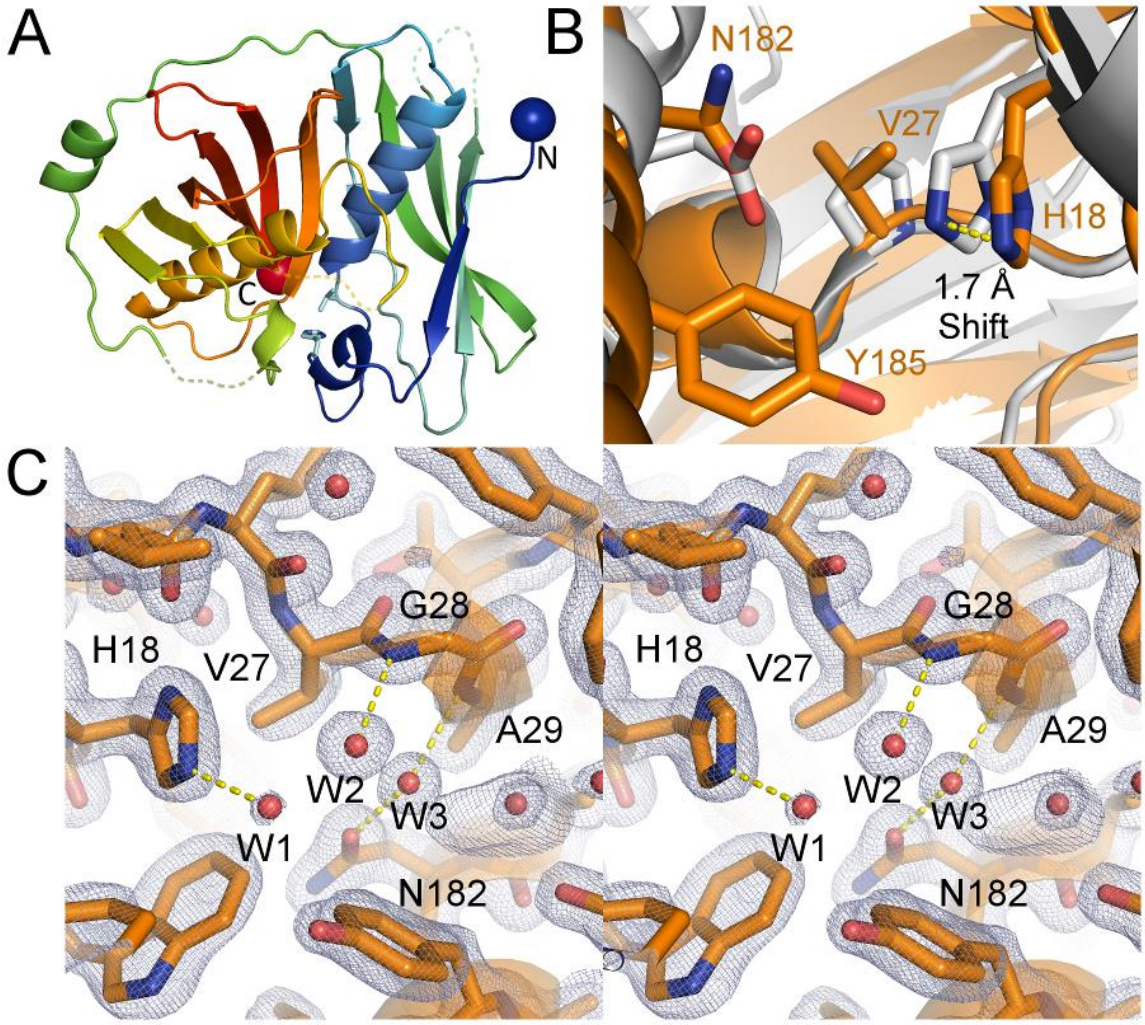
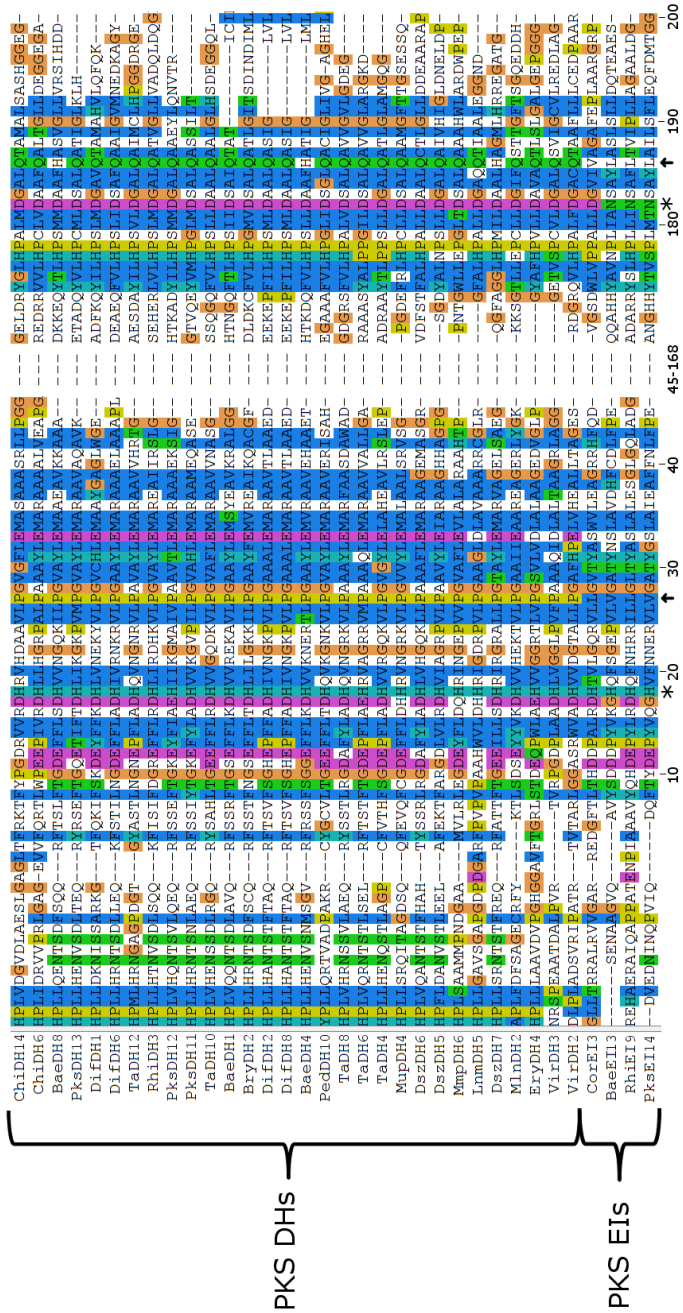


Figure 5-2. Structure of PksEI14. **(a)** The monomeric crystal structure of PksEI14 is shown in cartoon form, with N- and C-termini labeled with blue and red spheres, respectively. Flexible loops that could not be modeled into electron density are shown as dashed lines. The catalytic histidine (His18) and active site valine (Val27) are modeled as sticks. **(b)** A structural alignment of PksEI14 (orange) with the DH domain from the erythromycin PKS (gray, PDB Code 3EL6) reveals that the highly conserved proline in the PKS DH HX_8P motif is replaced with a valine, inducing a steric repositioning of the catalytic His18 imidazole ring. Asn182 replaces the catalytic aspartic acid of the PKS DH domain. The only active site residue that could complete a catalytic dyad is Tyr185; however, mutation to phenylalanine did not abolish isomerase activity. **(c)** Stereodiagram of active site $2F_o$-F_c electron density contoured at 1.5 Å r.m.s.d. Yellow dashes indicate hydrogen bonds to ordered waters (each is between 2.8 and 3.1 Å). See also Figure 5-3.



(Figure 5-3 continued on next page.)

Figure 5-3. PKS DH and EI sequence alignment. Residue numbering is based on the structure of PksEI14 (last in the alignment). Asterisks indicate the catalytic dyad of the DH domain, and arrows indicate residues that apparently alter the function of the DH catalytic dyad in EI domains. In the sequence motif HX_8P, His18 is responsible for deprotonating the α-carbon during dehydration, and the imidazole ring of this residue stacks against Pro27, highly conserved among DH sequences. Pro27 is substituted by either valine or leucine in EI sequences, which serves to shift the catalytic imidazole ring approximately 1.7 Å relative to its observed position in PKS DH domains. In the sequence motif DX_3Q/H, Asp182 acts as a general acid during dehydration and is activated by a hydrogen-bond network bridged by Gln/His186 (Gay *et al.*, 2013). The deterioration of this motif in the EI domain can be seen in the alignment, with Asp182 replaced with asparagine in three of the four EI domains and Gln/His186 replaced with leucine/valine. (Accession codes for sequences in the alignment can be found in the Supplementary Information for "A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase." Gay *et al.*, 2014).

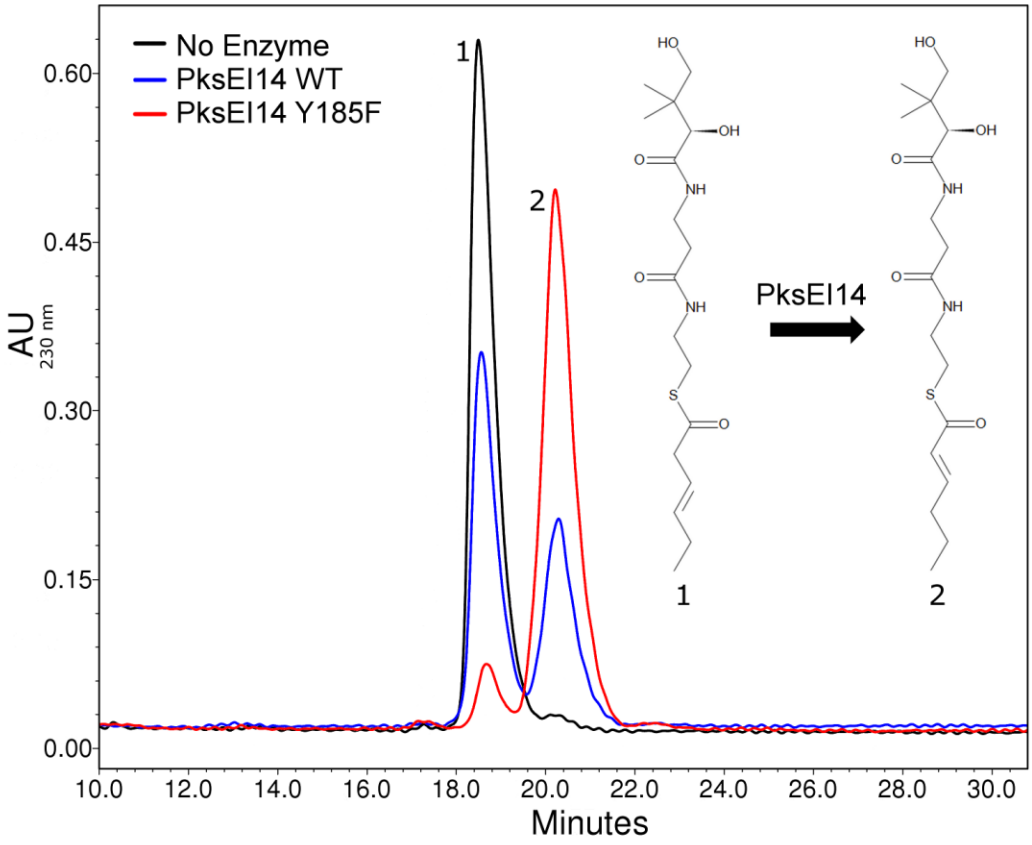
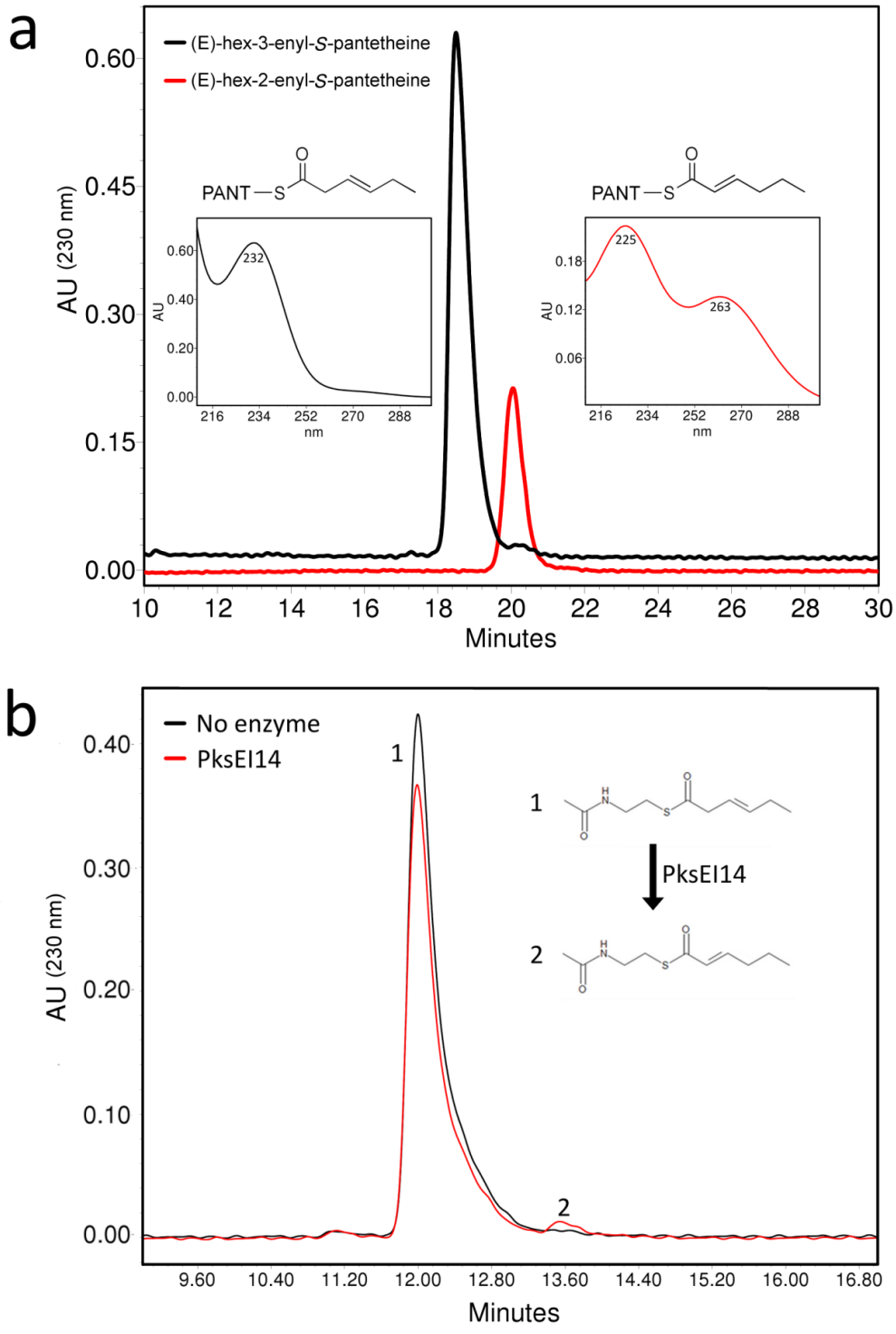


Figure 5-4. PksEI14 activity assay. Reversed-phase HPLC chromatograph of reconstituted *in vitro* isomerase activity observed after incubation of the β,γ-unsaturated substrate analogue with isolated PksEI14. The black line represents a control reaction that did not include any enzyme, revealing a single peak for (*E*)-hex-3-enyl-*S*-pantetheine. The blue line represents the same reaction with the addition of wild-type PksEI14, revealing a new peak with properties consistent with a synthetic standard of (*E*)-hex-2-enyl-*S*-pantetheine, the β,γ → α,β isomerized product. An identical reaction replacing wild-type PksEI14 with the Tyr185Phe mutant is shown in red, exhibiting increased isomerase activity over the same time course. See also Figure 5-5.



(Figure 5-5 continued on next page.)

Figure 5-5. Characteristics of synthetic substrate and product standards. **(a)** Reversed-phase HPLC analysis of the reaction of PksEI14 with (*E*)-hex-3-enyl-*S*-pantetheine revealed a new peak with a slightly delayed retention time (Figure 5-4). To confirm that this peak was the isomerized reaction product, synthetic (*E*)-hex-2-enyl-*S*-pantetheine was also analyzed by reversed-phase HPLC. Shown in black with a retention time of 18.5 minutes is synthetic (*E*)-hex-3-enyl-*S*-pantetheine, which is the same trace shown in Figure 5-4 (also in black). The maximal UV-absorbance for this β,γ-unsaturated thioester is 232 nm (left inset). The synthetic α,β-unsaturated thioester (*E*)-hex-2-enyl-*S*-pantetheine has a slightly delayed retention time of 20.1 minutes, and UV-absorbance maxima of 225 and 263 nm (shown in red, right inset), which match the PksEI14-catalyzed isomerization products. **(b)** To determine if PksEI14 was active towards NAC-tethered substrates, (*E*)-hex-3-enyl-*S*-NAC was incubated with PksEI14 under similar conditions to the assays described for pantetheine-tethered substrates. The isomerization activity observed was severely attenuated, indicating that the pantetheinyl-moiety is recognized by the enzyme. The identity of the reaction product, (*E*)-hex-2-enyl-*S*-NAC, was confirmed by the UV-absorbance maxima of the new peak at 13.6 minutes (225 and 263 nm).

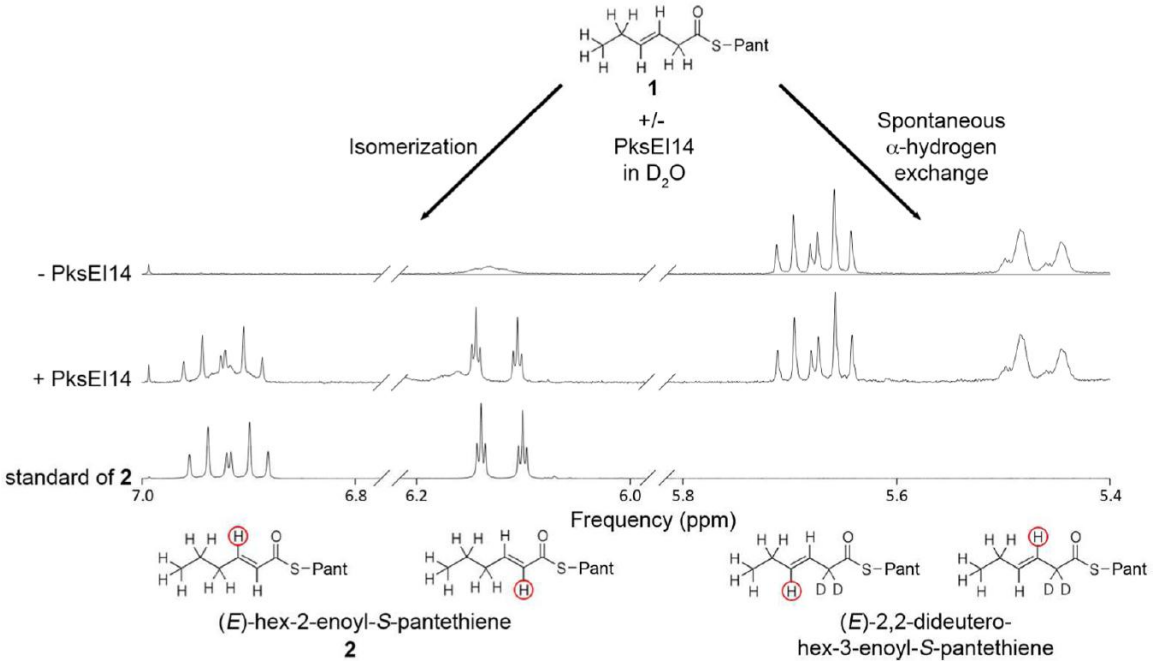
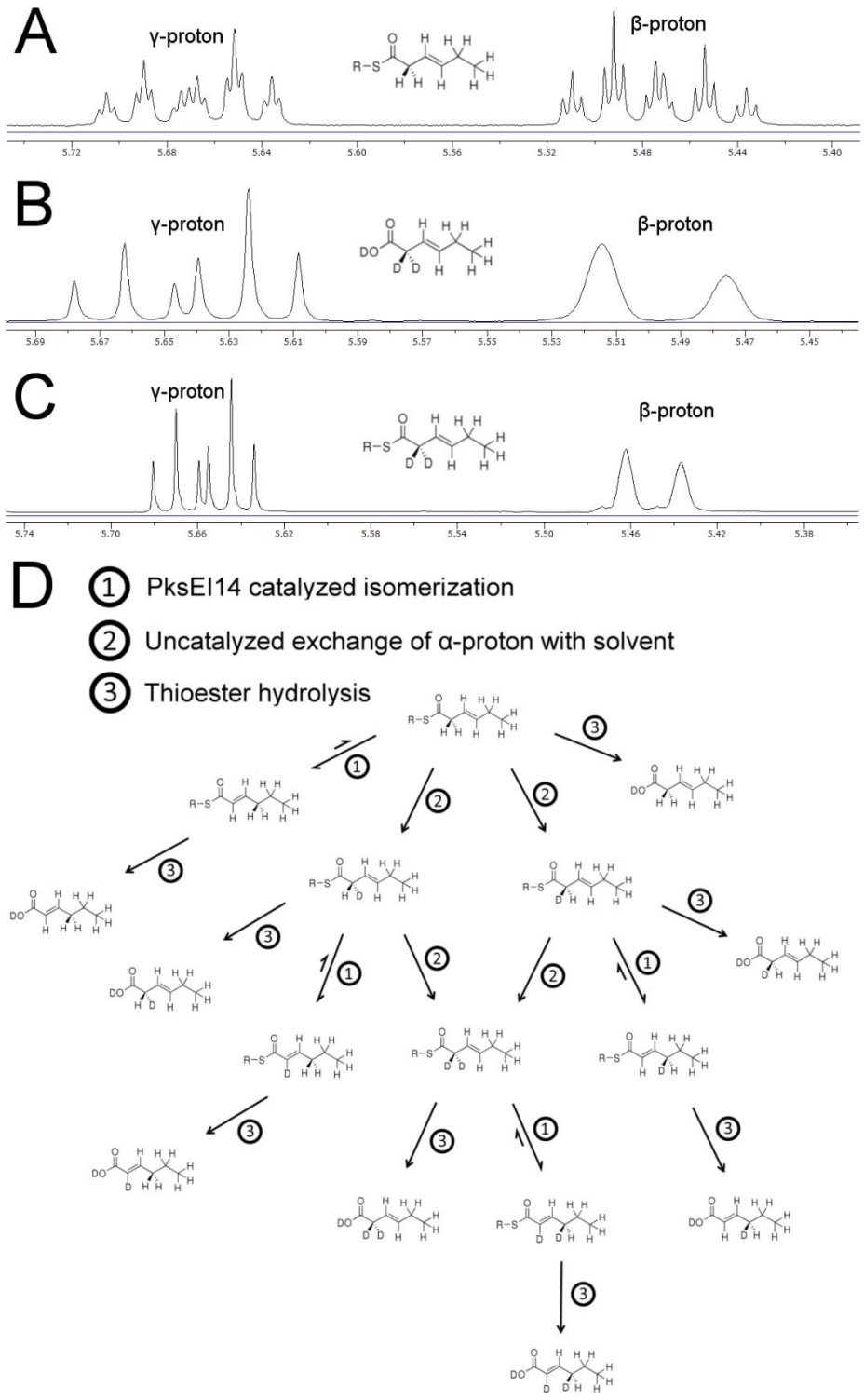


Figure 5-6. PksEI14-catalyzed isomerization in D2O. To determine whether PksEI14 shuttles a proton within the substrate or employs water in the catalytic mechanism, (*E*)-hex-3-enyl-*S*-pantetheine (1) was incubated with PksEI14 in D2O, and the reaction was analyzed by NMR. The negative control (top spectrum), in which PksEI14 was not added, shows only uncatalyzed α-hydrogen exchange. When PksEI14 is supplied (middle spectrum), isomerization to (*E*)-hex-2-enyl-*S*-pantetheine (2) is observed. If the mechanism of isomerization involved a solvent- or enzyme-donated proton, the γ-position would be monodeuterated, and different splitting patterns would be observed. The bottom spectrum is of synthetically prepared (*E*)-hex-2-enyl-*S*-pantetheine (2). Each of the four possible olefinic proton signals are aligned with the corresponding region of the spectrum and denoted with red circles. Pantetheine moieties are represented by "Pant". See also Figure 5-7.



(Figure 5-7 continued on next page.)

Figure 5-7. Background exchange of α-protons in β,γ-unsaturated thioesters. The solvent-catalyzed exchange of protons for deuterons at the α-position of β,γ-unsaturated thioesters complicates the analysis of isomerase reactions conducted in deuterated solvent. In each panel, "R" represents D-pantetheine. **(a)** The low-field region from a ^{1}H-NMR spectrum of (*E*)-hex-3-enyl-*S*-pantetheine reveals a multiplet splitting pattern for each olefinic proton. **(b)** The synthesis of (*E*)-2,2-dideutero-hex-3-enyl-*S*-pantetheine required the initial synthesis of (*E*)-2,2-dideutero-hex-3-enoic acid, and the low-field ^{1}H-NMR region of the α-deuterated free acid is shown. The multiplet olefinic signals are replaced by peaks with interpretable splitting patterns, which reveal a doublet of triplets for the γ-proton, and a broad doublet for the β-proton. **(c)** During the incubation of PksEI14 with (*E*)-hex-3-enyl-*S*-pantetheine in D_2O, continuous HPLC analysis revealed that the reaction had stalled, and it was therefore quenched and subjected to ^{1}H-NMR. Shown is the low-field region of the spectrum, which indicates that the (*E*)-hex-3-enyl-*S*-pantetheine substrate had nearly completely converted (uncatalyzed) to (*E*)-2,2-dideutero-hex-3-enyl-*S*-pantetheine. The β-proton signal for the isomerized reaction product is also visible in the same spectrum (shown in Figure 5-6), and reveals that the enzyme-catalyzed isomerization (which must have occurred before the substrate underwent uncatalyzed α-deuteration) yielded a γ-protonated product. **(d)** Due to the uncatalyzed background exchange of α-protons in the β,γ-unsaturated substrate mimic, analysis of PksEI14 reaction products becomes very complicated. Shown are all the competing reactions that occur when (*E*)-hex-3-enyl-*S*-pantetheine is incubated with PksEI14 in D_2O.

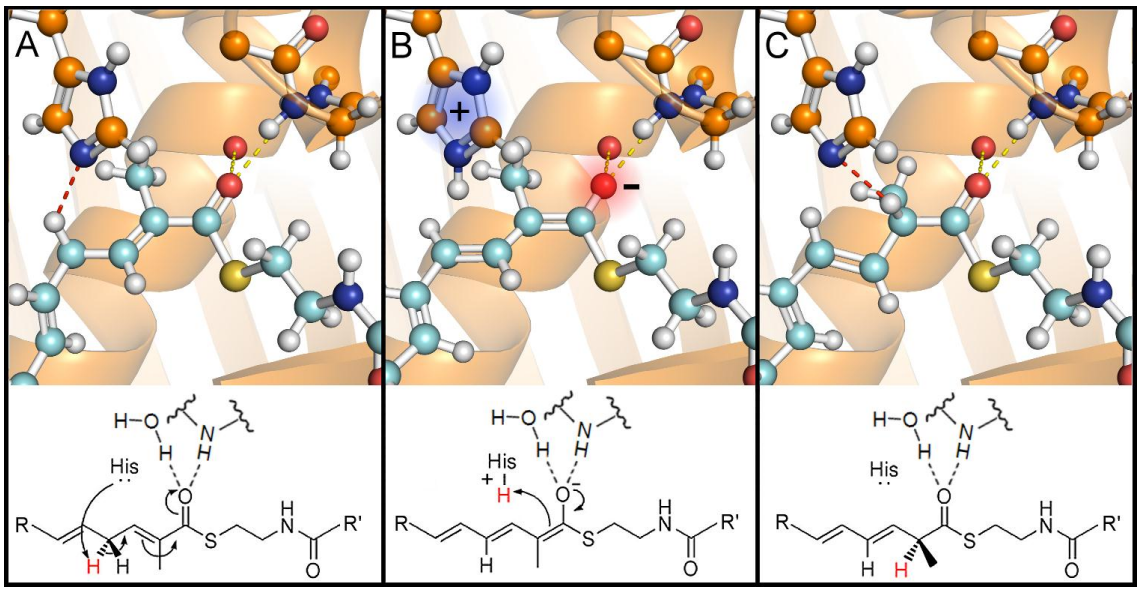


Figure 5-8. Proposed mechanism for double bond migration. Shown in three panels is the proposed mechanism for the biological reaction catalyzed by the PKS EI domain. Bacillaene carbon atoms are modeled in light blue, "R" represents bacillaene atoms not shown in the figure, and "R′" represents the ACP phosphopantetheinyl arm. Each yellow dashed line indicates a distance of 2.8 – 3.1 Å, and the red dashed lines indicate a distance of 2.2 Å. **(a)** The substrate modeled contains an α-methyl branch and is unsaturated at the α,β and δ,ε positions to resemble the biological intermediate. The thioester oxygen atom is modeled in the place of Wat2 (see Figure 5-2C). Catalysis is initiated by the abstraction of a γ-proton by the catalytic histidine. **(b)** The enolate is stabilized by the polarization of the thioester carbonyl, which forms interactions with the positive dipole of a central α-helix. **(c)** In the final panel, the proton abstracted from the γ-position has been shuttled to the α-position.

Data collection	Native	CH_3-Hg Derivative
Wavelength (Å)	1.0332	0.9763
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions, *a*, *b*, *c* (Å)	48.8, 68.3, 77.8	48.7, 68.2, 78.1
Resolution (Å)	35.37 – 1.73	51.3 – 1.88
R_{merge}	0.113 (0.702)	0.066 (0.520)
$I/\sigma(I)$	48.3 (2.5)	26.1 (5.1)
No. of reflections	25889 (1429)	21801 (3088)
Completeness (%)	97.6 (74.4)	99.5 (98.9)
Redundancy	3.5 (2.4)	14 (13.7)
Wilson B value (Å^2)	41.3	24.7
No. of heavy atom sites	-	2
Figure of merit	-	0.37
Refinement		
Resolution (Å)	35.37-1.73	
No. of reflections	25889 (1429)	
R_{work}/R_{free}	0.214/0.239	
No. of atoms		
Protein	1932	
Water	131	
Average *B* factors (Å^2)		
Protein	41.1	
Water	50.2	
RMS deviations		
Bond lengths (Å)	0.010	
Bond angles (°)	1.424	
Ramachandran Statistics (%)		
Preferred Regions	98.72	
Allowed Regions	1.28	
Outliers	0.0	

Table 5-1. Crystallographic data and refinement statistics. Values in parentheses refer to the highest resolution shell (1.76-1.73 Å).

Chapter 6. A close look at a ketosynthase from a *trans*-acyltransferase modular polyketide synthase

ABSTRACT

The recently discovered *trans*-acyltransferase modular polyketide synthases catalyze the biosynthesis of a wide range of bioactive natural products in bacteria. Here we report the first structure from a *trans*-acyltransferase polyketide module – the second ketosynthase of the bacillaene synthase. This 1.95 Å-resolution structure provides the highest resolution view available of a modular polyketide synthase ketosynthase and reveals a flanking subdomain that is homologous to an ordered linker in *cis*-acyltransferase modular polyketide synthases. The structure of the cysteine-to-serine mutant of the ketosynthase acylated by its natural substrate provides high-resolution details of how a native polyketide intermediate is bound and helps explain the basis of ketosynthase substrate specificity. The substrate range of the ketosynthase was further investigated by mass spectrometry.

INTRODUCTION

The complexity of reduced polyketides is a reflection of the polyketide synthases (PKSs) that produce them - a series of multidomain modules that catalyze the elongation and processing of polyketide intermediates in assembly-line fashion (Piel, 2010; Keatinge-Clay, 2012). While PKS modules are comprised of domains equivalent to those of type I fatty acid synthases (FASs), their products are much more functionalized due to the processive nature of these assembly lines and the combination of enzymes present within their component modules. Since invaluable human medicines such as the

Chapter 6 was adapted from the following published work: Gay DC, Gay G, Axelrod AJ, Jenner M, Kohlhaas C, Kampa A, Oldham NJ, Piel J, and Keatinge-Clay AT. (2014) A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. *Structure* **22**, 444–451. Each author contributed to the experimental design and manuscript preparation.

antibacterial erythromycin, the antifungal amphotericin, and the anticancer agent epothilone are produced by these biosynthetic machines, it is anticipated that understanding their architectures and activities will accelerate the development of new medicines.

Presently, all of the available structural information for PKS assembly lines is from *cis*-acyltransferase (*cis*-AT) systems (Keatinge-Clay, 2012), in which AT domains are integrated into the assembly line polypeptides C-terminal to ketosynthase (KS) domains. Within the modules of *cis*-AT PKSs, the AT domains are rigidly connected to the dimeric KS through an ~100-residue subdomain also present in metazoan FASs (Tang *et al.*, 2006; Tang *et al.*, 2007; Maier *et al.*, 2008; Pappenberger *et al.*, 2010). In contrast, the AT domains of *trans*-AT PKSs are not integrated into the assembly lines but expressed as free-standing polypeptides (Patel *et al.*, 1995; Calderone *et al.*, 2006; Chen *et al.*, 2006, Butcher *et al.*, 2007; Moldenhauer *et al.*, 2007; Reddick *et al.*, 2007; Moldenhauer *et al.*, 2010; Jenner *et al.*, 2013) (**Figure 6-1**). These discrete ATs usually transfer malonyl extender units to several acyl carrier protein (ACP) domains within a synthase (Cheng *et al.*, 2003); however, some transfer more unusual extender units to specific ACPs (Musiol *et al.*, 2011). The first reported *trans*-AT PKS, which has become prototypical for this PKS class, is the bacillaene synthase found within *Bacillus subtilis* and *Bacillus amyloliquefaciens* (*pksX* and *baeX* gene clusters, respectively) (Piel *et al.*, 2010). Bacillaene (**Figure 6-1**) is a linear polyene diamide that inhibits prokaryotic protein synthesis (Patel *et al.*, 1995).

The carbon-carbon bond-forming chemistry performed by KSs is at the core of polyketide synthesis (**Figure**s **6-1** and **6-2**). Catalysis is initiated when a polyketide intermediate is transferred through transthioesterification from the ACP of one module to the KS of the subsequent module. An extender unit-bound ACP then docks to the polyketide-bound KS to enable a decarboxylative condensation, generating a β-keto intermediate that may be processed at its α- and β-carbons through the action of a methyltransferase (MT), ketoreductase (KR), dehydratase (DH), enoylreductase (ER), or β-branching enzyme. Unlike KSs from *cis*-AT PKSs, KSs from *trans*-AT PKSs cluster into distinct clades that correlate with the chemistries of the α- and β-positions of the accepted polyketide intermediate (Nguyen *et al.*, 2008). Recently it was shown that, at least in some cases, this sequence-based correlation manifests in pronounced substrate specificity (Jenner *et al.*, 2013; Kohlhaas *et al.*, 2013). An understanding of substrate specificity at the structural level is therefore crucial for successful engineering of *trans*-AT PKSs.

Here we report the 1.95 Å-resolution crystal structure of the KS from the second PKS module of the bacillaene synthase (PksKS2), the highest resolution available to date of a KS domain from a modular PKS. The structure reveals a flanking subdomain of the KS that is equivalent to the KS/AT adapter of *cis*-AT PKSs and contains two loosely-connected helices where the AT is incorporated in *cis*-AT systems. The structure of a cysteine-to-serine mutant acylated by both a hexanoyl group and the natural polyketide intermediate help elucidate how *trans*-AT KSs control substrate specificity, identifying the residues most likely to impart this substrate tolerance. These findings were further

elaborated through a mass spectrometry-based approach recently developed to investigate the substrate specificity of *trans*-AT KSs (Jenner *et al.*, 2013).

RESULTS

Structure determination

The boundaries chosen for PksKS2 were based on those used to express KS+AT fragments from *cis*-AT PKSs, beginning six residues N-terminal to the IAIIG motif and ending six residues C-terminal to the LPxYPFxxxxxW motif (Tang *et al.*, 2006; Tang *et al.*, 2007). The 1.95-Å structure of PksKS2 was solved by molecular replacement using a KS from the erythromycin synthase (PDB code: 2QO3) as a search model (Tang *et al.*, 2007), revealing the KS dimer interface anticipated from related enzymes (**Figure**s **6-3** and **6-4**). Crystal structures of PksKS2(Cys176Ser) bound to the natural substrate and a hexanoyl substrate were obtained through soaking and co-crystallization, respectively, of substrate analogs (**Figure 6-5**).

Comparison of PksKS2 to other type I synthases

The architecture of the PksKS2 dimer is highly homologous to the equivalent regions of other type I synthases (C_α r.m.s.d.: 1.7 Å over 470 C_α with third module of the erythromycin PKS and 1.2 Å over 350 C_α with the human FAS) (Tang *et al.*, 2007; Pappenberger *et al.*, 2010). At the C-terminal end of the ~450-residue KS domain is incorporated an ~100-residue subdomain with a β-α-α-β-α-β-α fold (here referred to as the "flanking subdomain" and sometimes referred to as the "KS/AT adapter" in *cis*-AT

PKSs or "ATd" in *trans*-AT PKSs) (**Figure**s **6-3**, **6-4**, and **6-6**) (Keatinge-Clay, 2012; Calderone *et al.*, 2006). In both *cis*-AT PKSs and the metazoan FAS, the AT domain inserts between the final β-strand and α-helix of the ~100-residue subdomain, while in PksKS2, two loosely-connected helices, α17 and α18, insert at this position. The final loop of the KS in *trans*-AT PKSs is essentially identical to that in *cis*-AT PKSs, with an extended LPxYPF peptide mediating the interaction with the flanking subdomain. The equivalent peptide in metazoan FASs forms a helical turn. In all type I systems, a conserved tryptophan at the end of the final KS loop buries into a pocket near the two-fold axis of the synthase.

Compared to the metazoan FAS KS, the KSs of *trans*-AT and *cis*-AT PKSs are structurally more similar to one another. Only seven residues were found to consistently differ between *trans*-AT and *cis*-AT PKSs (**Table 6-2**). Four of these are within 15 Å of one another and together stabilize the position of helix α9. Although oriented by distinct networks of residues, this three-turn helix on the KS surface is equivalently located in *cis*-AT and *trans*-AT PKSs and absent from the KSs of metazoan FASs. Both types of PKS KS possess a longer, structured loop between α1 and α2 compared to the KSs of metazoan FASs.

KS specificity

The active site of PksKS2 is homologous to the active sites of the KSs from *cis*-AT PKS and metazoan FASs, with catalysis mediated by a cysteine and two histidines located within Tx**C**SSS, **H**GTGT, and KSNIG**H** motifs (as in *cis*-AT PKSs) (**Figure**s **6-3**

and **6-6**) (von Wettstein-Knowles *et al.*, 2006; Zhang *et al.*, 2006). Phylogenetic studies of KS domains from *trans*-AT systems have revealed that, in contrast to the KS domains from *cis*-AT systems, a correlation exists between the biosynthetic intermediate that a KS accepts and the primary sequence of that KS (Nguyen *et al.*, 2008). All residues within 15 Å of the active site cysteine were examined for each of the sixteen defined KS clades since sets of these residues are likely responsible for imparting substrate specificity (**Figure 6-5**).

Crystal structures of PksKS2 bound to substrates were sought in order to help elucidate the physical interactions of polyketide intermediates with KS active sites (**Figure**s **6-5** and **6-7**). A complex structure could not be obtained using crystals of wild-type PksKS2. However, by mutating the active site cysteine to serine and incubating PksKS2(Cys176Ser) with (*S*)-α-hydroxyisocaproyl-γ-aminobutyryl-*S*-*N*-acetylcysteamine (NAC), the natural intermediate was observed covalently bound to the active site serine (Calderone *et al.*, 2008; Olsen *et al.*, 2001). The ester carbonyl forms a hydrogen bond with the Ile433 NH, and a well-ordered water molecule adjacent to the ester (Wat1) is coordinated by the catalytic histidines. In unbound PksKS2, this water is only forming a hydrogen bond to His311; a similarly-located water has been proposed to participate in the condensation reaction (**Figure 6-3c**) (Zhang *et al.*, 2006). Several functional groups of the natural intermediate make favorable interactions with active site residues: its amide NH forms a hydrogen bond with the hydroxyl group of Tyr215, its hydroxyl group forms a hydrogen bond with the carboxylate of Glu362, and its gem-dimethyl terminus makes hydrophobic contacts with Leu363 and Ile151'. By co-crystallizing PksKS2(Cys176Ser)

with the substrate mimic hexanoyl-*S*-pantetheine, a hexanoyl group was also observed in the active site. As with the natural intermediate, the ester carbonyl forms a hydrogen bond with the NH of Ile433, and a well-ordered water molecule is coordinated by the catalytic histidines. The α- and β-carbons of the natural substrate and the hexanoyl group are closest to the side chains of Tyr237, Glu362, and Ile433. The dihedral angles neighboring the α- and β-carbons of the hexanoyl group and the natural intermediate slightly differ due to favorable contacts made between the natural intermediate and active site residues. No electron density is observed for either the NAC or pantetheine substrate handles, indicating that transesterification liberated the handles into solvent. Mass spectrometry suggests complete acylation of PksKS2(Cys176Ser) by the natural substrate.

A time-dependent, mass spectrometry-based method was recently developed to examine KS specificity (Jenner *et al.*, 2013). This method was utilized to examine the specificity of BaeKS2 with acyl-*S*-NAC substrates (**Figure**s **6-8** and **6-9**). This KS from the Bacillus amyloliquefaciens bacillaene synthase is the equivalent of PksKS2 and possesses identical residues within 15 Å of the catalytic cysteine. BaeKS2 was most rapidly acylated by the short β-ketoacyl-*S*-NAC **2**. The acetamide analog of the natural substrate, **3**, and the other unbranched acyl-*S*-NACs **1**, **4**, and **6** were also accepted. BaeKS2 tolerated a methoxy group in the δ- position but not the β-position, as indicated by the acceptance of **5** but not **7**.

The flanking subdomain

Most KSs from *trans*-AT PKSs, including those from clades III and X that do not catalyze chain elongation, contain a flanking subdomain (**Figure 6-10**). Only KSs from clades XIV and XV, located within so-called bimodules, consistently do not contain this subdomain (Piel, 2010; Nguyen *et al.*, 2008). Some flanking subdomains contain an inactive ACP domain, inserted between α14 and α15, lacking the serine residue to which the phosphopantetheinyl arm is normally attached (**Figure 6-11**).

The interaction of the AT domain with the KS/AT adapter in *cis*-AT PKSs is primarily mediated by hydrophobic interactions between an α-helix structurally equivalent to α19 of the flanking subdomain and a groove in the AT domain. While the C-terminal helix of some acyltransferases (e.g., *Streptomyces coelicolor* FabD) is located in this groove, discrete ATs from *trans*-AT PKSs often do not possess this helix (Keatinge-Clay *et al.*, 2003; Wong *et al.*, 2011). However, docking an AT to PksKS2 through α19 results in a clash with α17 and α18 of the flanking subdomain (**Figure 6-12**) (Tang *et al.*, 2007).

DISCUSSION

KS specificity is a central feature in *trans*-AT PKSs. The primary role of many KSs is gatekeeping – they catalyze the specific transfer of intermediates through the assembly line but do not catalyze chain elongation. One example from the bacillaene synthase is its final KS, which is condensation-incompetent but apparently ensures that a polyketide intermediate has been properly isomerized by the previous module before

allowing it to be off-loaded from the assembly line (**Figure 6-2**). Condensation-competent KSs also perform gatekeeping roles, as KSs of the same clade generally possess similar specificities (Jenner *et al.*, 2013) (**Figure 6-5** and **6-6**). PksKS2 belongs to Clade V, members of which usually accept α,β-unsubstituted substrates and, based on the presented crystallographic data, could employ Tyr237 and Glu362 to select against α- and β-branched substrates (β-keto substrate **2** may be able to acylate BaeKS2 since it is shorter than the natural substrate and possesses more degrees of freedom in the active site) (Nguyen *et al.*, 2008) (**Figure**s **6-7** and **6-13**). Clade I KSs possess smaller residues in these positions (phenylalanine and alanine, respectively) that likely provide more space for the variety of branched substrates they accept. The presented KS structures suggest that the methionine N-terminal to the catalytic cysteine in α,β-olefin-accepting Clade IX KSs sterically blocks α,β-branched and α,β-saturated substrates, which possess greater steric bulk around their α- and β-carbons than α,β-olefin substrates (Jenner *et al.*, 2013). The asparagine N-terminal to the catalytic cysteine in Clade XVI KSs plays a critical role in the acylation of KSs by amide-containing substrates generated by non-ribosomal peptide synthetase modules (Kohlhaas *et al.*, 2013). Modeling an asparagine residue in the place of Ala175 in PksKS2 reveals that its side chain carbonyl is in position to form a hydrogen bond with the amide NH of those intermediates. While it is apparent that threading the sequence of a KS through the PksKS2 structure will be beneficial in modeling interactions between that KS and its substrate, to determine the precise contacts formed the crystal structure of that acylated KS must be obtained. As demonstrated,

mutating the catalytic cysteine to serine prior to acylation with a small molecule analogue yields a highly stable ester adduct, enabling the visualization and analysis of these informative complexes.

Whether docking sites for discrete ATs exist for *trans*-AT PKS KSs is unknown. No feature of KS correlates with AT docking since no consistent differences are apparent between the surfaces of *trans*-AT KSs and *cis*-AT KSs. The only consistent difference between the flanking subdomain from *trans*-AT KSs and the equivalent region from other type I systems are the two inserted helices (α17 and α18) instead of an AT domain. A discrete AT could dock to α19 of PksKS2, through an interaction similar to EryAT3 with the equivalent helix of the KS/AT adapter. Such an interaction could be made possible by discrete ATs (e.g., PksC in the bacillaene synthase) through the absence of the C-terminal helix observed in *Streptomyces coelicolor* FabD (Keatinge-Clay *et al.*, 2003). However, α17 and α18 of the flanking subdomain would need to shift to enable the AT to dock in this orientation (**Figure 6-12**). The length and composition of α17 and α18 are not highly conserved (**Figure 6-10**).

Many condensation-incompetent KSs contain flanking subdomains that would serve no purpose if their only role was to bind AT. Since such a KS must still interact with two ACPs to transfer a polyketide through the assembly line, one could hypothesize that the flanking subdomains help form the ACP docking sites. However, even if the flanking subdomain facilitates interactions with ACPs, their involvement is not always required. KSs in clade XIV catalyze polyketide transfer between two ACPs without this subdomain, and KSs in clade XV are able to perform chain elongation without this

subdomain. In the bacterial FAS, the KS equivalents do not possess the subdomain but are docked by carrier proteins for transthioesterification and chain elongation reactions.

The 1.95-Å structure of PksKS2 reported here represents the first structure from a *trans*-AT PKS module, provides the highest resolution view of a type I KS domain, and enables a first glimpse of the mysterious flanking subdomain. The crystallographic visualization of a KS bound to its natural intermediate signifies the first snapshot of a polyketide intermediate covalently attached to an enzyme from within a PKS module and reveals the locations of residues that control substrate specificity. The structure of the KS active site correlates with phylogenetic studies of KS specificity that were further elucidated in this work by mass spectrometry. Together, the reported structural and functional analysis provides much-needed, high-resolution insights into the most central enzyme of PKS assembly lines.

MATERIALS AND METHODS

Cloning, expression, and purification of PksKS2

Domain boundaries for the polypeptide fragment containing PksKS2 were chosen based on the crystal structure of Ery(KS+AT)3, six residues N-terminal of the conserved IAIXG motif and six residues C-terminal to a LPxYPFxxxxxW motif (Tang *et al.*, 2007). The DNA encoding PksKS2 was amplified using primers 5'-GCGGCCTGGTGCCGCGCGGCTCTAGCGCTGCCGATTTTGAGCCG-3' and 5'-GTGGTGGTGGTGGTGGTGATGTTAATGTCCCTTTGCCGGTACCC-3' (ligation-independent cloning regions underlined) from *Bacillus subtilis subsp*. 168 and inserted

into pGAY28b, a ligation-independent cloning vector constructed from pET28b. *E. coli* BL21(DE3) transformed with the expression plasmid was inoculated into LB media containing 50 mg/L kanamycin at 37 °C, grown to OD_{600} = 0.4, and induced with 0.5 mM IPTG. After 12 h at 15 °C, cells were collected by centrifugation and resuspended in lysis buffer (0.5 M NaCl, 10% (v/v) glycerol, 0.1 M HEPES, pH 7.5). Following sonication, cell debris was removed by centrifugation (30,000 rcf for 30 min.). The supernatant was poured over a column of Nickel-NTA resin (Thermoscientific), which was then washed with 50 mL lysis buffer containing 15 mM imidazole and eluted with 5 mL lysis buffer containing 150 mM imidazole. PksKS2 was further purified using a Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated with 150 mM NaCl, 10 mM HEPES, pH 7.5. The eluted protein was concentrated to 9 mg/mL in the equilibration buffer and stored at -80 ˚C until needed.

The PksKS2(Cys176Ser) expression plasmid was generated by PCR amplification from the PksKS2 construct described above using primers 5'-GAGACGGCCTCGAGCAGTTCTCTTGTAGCCATTCACCGTGCTG-3' and 5'-AGAACTGCTCGAGGCCGTCTCTACTGGCTCGCTCGGACCATGAATA-3' (XhoI site underlined). The PCR product was digested with XhoI and ligated. Purification of the mutant protein was performed as described for unmutated PksKS2.

Crystallization and structure determination

Crystals of PksKS2 grew in 2-14 d by sitting drop vapor diffusion at 22 °C. Drops were formed by mixing 2 µL protein solution (7 mg/mL PksKS2, 150 mM NaCl, 10 mM

HEPES, pH 7.5) with 1 µL crystallization buffer (1.1 M ammonium sulfate, 0.2 M lithium sulfate, 0.1 M Tris, pH 8.0). Crystals were frozen in liquid nitrogen after a 20-min soak in the crystallization buffer modified with 20% (v/v) glycerol. Diffraction data, collected at ALS Beamline 5.0.3, were processed by HKL2000 (Otwinowski & Minor, 1997). The structure was solved to 1.90 Å resolution by molecular replacement with PhaserMR (McCoy *et al.*, 2007) in the CCP4 suite (Winn *et al.*, 2011), using a poly-alanine model (Stein, 2008) of the KS monomer from Ery(KS+AT)3 (PDB code: 2QO3) as the search model. The model generated from the molecular replacement solution was used to iteratively build into the remaining electron density map with Coot (Emsley *et al.*, 2010) and was refined with Refmac5 (Murshudov *et al.*, 1997).

Crystals of PksKS2(Cys176Ser) bound to a hexanoyl group grew over 2-6 d by sitting drop vapor diffusion at 22 °C. Drops were formed by mixing 2 µL protein solution (7 mg/mL PksKS2(Cys176Ser), 150 mM NaCl, 10 mM HEPES, pH 7.5, 10 mM hexanoyl-*S*-pantetheine) with 1 µL crystallization buffer (1.3 M ammonium sulfate, 0.1 M Tris, pH 8.2). Crystals were frozen in liquid nitrogen after a 20-min soak in crystallization buffer modified with 20% (v/v) glycerol and 10 mM hexanoyl-*S*-pantetheine, and the diffraction data were collected at ALS Beamline 5.0.2. The crystals were found to have the same space group and unit cell dimensions as those of unmutated PksKS2 and were iteratively refined with Coot and Refmac5.

Crystals of PksKS2(Cys176Ser) bound to the natural intermediate could not be optimized to diffract beyond 3.6-Å resolution under co-crystallization conditions. Thus a soaking strategy was employed to obtain a higher-resolution complex structure. Crystals

of PksKS2(Cys176Ser) grew over 1-2 d by sitting drop vapor diffusion at 22 °C in the absence of substrate. Drops were formed by mixing 1.5 µL protein solution (7 mg/mL PksKS2(Cys176Ser), 150 mM NaCl, 10 mM HEPES, pH 7.5) with 1 µL crystallization buffer (15% (w/v) PEG 3350, 0.18 M tri-ammonium citrate). Crystals were soaked for 2 h in crystallization buffer modified with 20% (v/v) glycerol and 10 mM (*S*)-α-hydroxyisocaproyl-γ-aminobutyryl-*S*-*N*-acetylcysteamine before being frozen in liquid nitrogen. Diffraction data were collected at ALS Beamline 5.0.2, and the structure was solved by molecular replacement with PhaserMR using an unmutated PksKS2 monomer as the search model. The structure was iteratively refined with Coot and Refmac5.

Cloning, expression, and purification of BaeKS2

The *Bacillus amyloliquefaciens* FZB42 DNA that encodes BaeKS2 was amplified by PCR with Phusion High-Fidelity DNA polymerase (NEB) using the forward primer 5´-AAAGGATCCGCAAAAGAGCATCCGGGCCGTTT-3´ and the reverse primer 5'-AAAAAAAAGCGGCCGCTCACCATTGTTTCGTCAGAATACGTTTCATCCG-3´. The restriction digest of the PCR amplicon was performed with *Bam*HI and *Not*I, and the product was purified on a TAE agarose gel. The insert was ligated into the *Bam*HI and *Not*I sites of pHis8 (Jez *et al.,* 2000). The ligation product was then transformed into electrocompetent *E. coli* BL21(DE3) cells. Positive clones were grown overnight in LB medium at 37 °C. The plasmid DNA was isolated and sequenced. A positive colony was used for gene expression.

A 10 mL overnight culture of *E. coli* BL21(DE3) transformed with the BaeKS2 plasmid (pAK6) were transferred into 1 L LB medium with 50 μg/mL kanamycin. After 3-4 h incubation at 37 °C and growth to OD_{600} = 0.7, overexpression was induced by addition of 1 mM IPTG, and the culture was grown overnight at 16 °C. Cells were centrifuged, and the supernatant was removed. The cell material was resuspended in lysis buffer (25 mM Tris, 0.5 M NaCl, 10% (v/v) glycerol, 10 mM imidazole, pH 7.6), sonicated, and centrifuged again. BaeKS2 was purified via its N-terminal His_8-tag by adding Ni-NTA agarose to the supernatant and incubating 1 h on ice while shaking. The suspension was loaded on a column and washed with lysis buffer with an increasing imidazole gradient (40-300 mM). Fractions containing BaeKS2 were frozen in liquid nitrogen and stored at -80 °C.

Mass spectrometry analysis of KS acylation

Acylation assays. Acylation reactions were conducted in storage buffer (25 mM Tris, 0.5 M NaCl, 10% (v/v) glycerol, pH 7.6). NAC thioesters were incubated with KS domains at a final concentration of 2 mM. BaeKS2 was maintained at 12 μM for acylation assays. To ensure the NAC thioesters remained in solution, the concentration of DMSO was adjusted to 5% (v/v). Acylation reactions were allowed to run 0-20 min at 25 °C before quenching with 0.1% TFA.

Sample preparation for mass spectrometry analysis. C_4 ZipTips™ (Merck Millipore, Billerica, MA) were used to prepare each sample; the C_4 ZipTip was washed

with two 10 μL aspirations of H_2O/0.1% TFA solution. The protein sample was then loaded onto the ZipTip column by 15-20 10-μL aspirations of the sample, each time injecting back into the sample vial. The loaded protein sample was then desalted by 15x10-μL aspirations of H_2O/5% MeOH/0.1% TFA, followed by elution of the sample into 5 μL H_2O/80% MeCN/0.1% TFA.

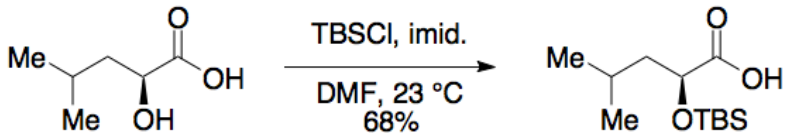
Data analysis. Denatured KS spectra obtained from acylation experiments were subjected to minimal smoothing and noise reduction. The spectra were deconvoluted using the transform function of Mass Lynx™, taking an average of all charge states. Relevant peak intensities were recorded and converted into the concentration of unacylated KS and acylated KS. Data was recorded in triplicate, and an average taken for kinetic plots.

Instrumentation. All experiments described within this study were analyzed using a SYNAPT™HDMS™, a hybrid quadrupole-ion mobility-orthogonal acceleration TOF instrument (oa-TOF) produced by Waters™ (Milford, MA). This instrument has multiple ionization capabilities of which nanoESI was utilized during this study. Parameters employed: cone voltage, 20 V; capillary voltage, 1.5 kV; trap CE, 10 V; transfer CE, 5 V; backing pressure, ~4 mbar; trap pressure, 2.3×10^{-3}; TOF pressure: 1.6×10^{-6} mbar. A quadrupole profile was applied that gave optimal transmission between 1000-3000 m/z. Spectra were acquired between 1000-4000 m/z. Mass Lynx™ 4.1 software was used to control the parameters of the SYNAPT and to view and analyze the raw data.

Mass spectrometry analysis of PksKS2 acylation

Acylation assays. Acylation reactions were conducted in storage buffer (10 mM HEPES, 0.15 M NaCl, pH 7.5). The NAC thioester of the natural substrate was incubated with Cys176Ser PksKS2 at a final concentration of 10 mM, while the protein was maintained at 100 μM. To ensure the NAC thioester remained in solution, the concentration of DMSO was adjusted to 0.5% (v/v). The reaction was allowed to run for 60 min at 25 °C before quenching with 0.1% TFA. Protein molecular weight determination was provided by the University of Texas at Austin Proteomics Facility on the 4000 QTrap (AB Sciex)**.**

Synthesis of ligands for crystallographic analysis of KS acylation



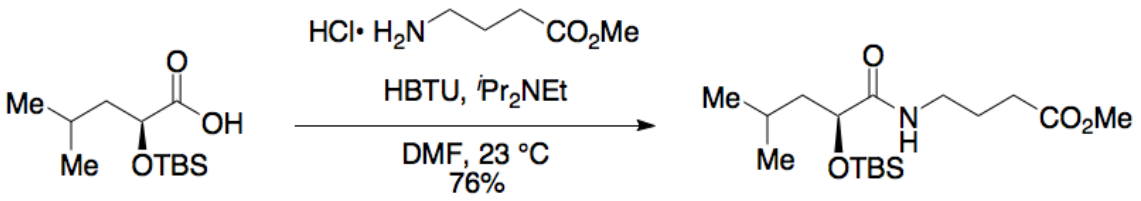
To a stirred solution of (*S*)-(-)-2-hydroxyisocaproic acid (500 mg, 3.8 mmol, 1.0 eq.) in DMF (2.0 mL) at 23 °C was added imidazole (1.24 g, 18.2 mmol, 4.8 eq.) followed by TBSCl (1.37 g, 9.1 mmol, 2.4 eq.). The pale yellow solution was stirred for 18 h and then diluted with 100 mL of 1:1 hexanes:EtOAc. The solution was washed with 10% aqueous citric acid (30 mL), saturated $NaHCO_3$ (50 mL), water (50 mL) and then brine (50 mL). The organic layer was dried over sodium sulfate and concentrated *in vacuo*. The residue

was dissolved in MeOH (40 mL), cooled to 0 °C and both K_2CO_3 (1.3 g, 9.5 mmol, 2.5 eq.) and water (12 mL) were added. The resulting mixture was stirred 4 h at 23 °C, and the solvent was removed *in vacuo*. The residue was dissolved in water (20 mL), cooled to 0 °C, and acidified to pH = 4 through the addition of 10% aqueous citric acid solution. The aqueous solution was washed with EtOAc (50 mL), and the organic layer was separated. The aqueous layer was washed with EtOAc (50 mL), and the combined organic layers were dried over sodium sulfate and concentrated *in vacuo*. The crude product was chromatographed on silica gel eluting with 2:1 EtOAc:hexanes to afford the title compound (68%, 640 mg) as a pale, clear oil.

^{1}H NMR: 400 MHz ($CDCl_3$) δ = 4.27 (dd, *J* = 7.2 Hz, *J* = 5.1 Hz, 1H), 1.82 (m, 1H), 1.67-1.56 (m, 2H), 0.95 (d, *J* = 6.6 Hz, 3H), 0.94 (s, 9H), 0.94 (d, *J* = 6.6 Hz, 3H), 0.14 (s, 3H), 0.13 (s, 3H).

^{13}C NMR: 100 MHz ($CDCl_3$) δ = 179.8, 70.4, 44.0, 25.7, 24.0, 23.2, 21.7, 18.2, -4.9, -5.3.

Spectroscopic values agree with the literature (Menche *et al.*, 2009).



To a stirred solution of protected carboxylic acid (140 mg, 0.57 mmol, 1.0 eq.) in DMF (2.9 mL) at 23 °C was added *N*,*N*-diisopropylethylamine (0.36 mL, 2 mmol, 3.5 eq.) followed by HBTU (259 mg, 0.68 mmol, 1.2 eq.). After 5 min, γ-aminobutyric acid methylester hydrochloride (107 mg, 0.68 mol, 1.2 eq.) was added in one portion. After 12 h, the reaction was diluted with 1:1 hexanes:EtOAc (100 mL) and washed with saturated aq. LiCl solution (3 x 50 mL). The organic layer was dried over sodium sulfate and concentrated *in vacuo* to furnish the title compound (150 mg, 76%) as a tan oil that did not require further purification. For analytical purposes, the material was chromatographed on silica gel eluting with 1:1 hexanes:EtOAc.

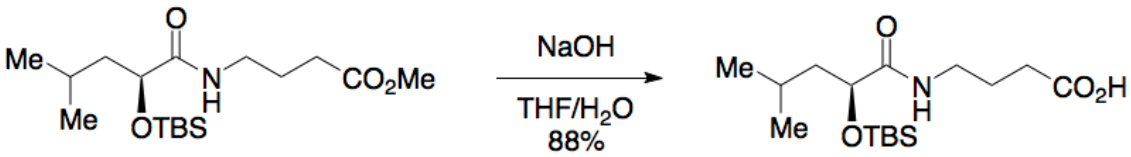
R_f = 0.5 (2:1 hexanes:EtOAc)

^{1}H NMR: 400 MHz ($CDCl_3$) δ = 6.64 (bs, 1H), 4.12 (dd, *J* = 7.2 Hz, *J* = 4.8 Hz, 1H), 3.68 (s, 3H), 3.29 (qd, *J* = 7.2 Hz, *J* = 3.1 Hz, 2H) 2.36 (t, *J* = 7.5 Hz, *J* = 7.2 Hz, 2H), 1.88-1.74 (m, 3H), 1.56 (m, 2H), 0.93 (s, 9H), 0.92 (d, *J* = 6.5 Hz, 3H), 0.91 (d, *J* = 6.8 Hz, 3H), 0.09 (s, 3H), 0.06 (s, 3H).

^{13}C NMR: 100 MHz ($CDCl_3$) δ = 177.2, 175.2, 72.3, 44.7, 38.1, 31.3, 25.7, 24.8, 23.9, 23.4, 22.2, 17.9, -4.9, -5.2.

IR: (film, υ cm^{-1}) 3426, 2969, 1745, 1675, 1519.

HRMS: ESI calc. for $C_{17}H_{36}NO_4Si$ $[M+H]^+$:346.2414. Found 368.2226 $[M+Na]^+$.



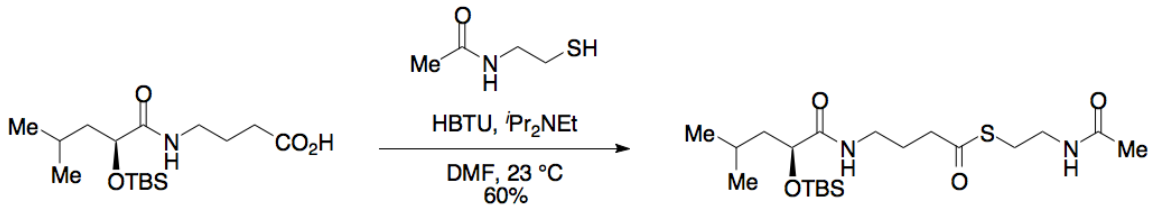
To a stirred solution of the methyl ester (150 mg, 0.43 mmol, 1.0 eq.) in THF (2.1 mL) at 0 °C was added a solution of aqueous NaOH (1.0 M, 1.29 mL, 1.29 mmol, 2.0 eq.), and upon complete addition the reaction was allowed to warm to 23 °C. At 5 h, the reaction was cooled to 0 °C and was treated with 1.0 M HCl (aqueous) (1.4 mL). The reaction was diluted with EtOAc (100 mL), and the aqueous layer was separated. The organic layer was dried over sodium sulfate and concentrated *in vacuo* to furnish the title compound (125 mg, 88%) as a white soft solid. The crude material was used without further purification.

^{1}H NMR: 400 MHz ($CDCl_3$) δ = 6.74 (bs, 1H), 4.15 (dd, *J* = 7.2 Hz, *J* = 4.8 Hz, 1H), 3.35 (m, 2H) 2.39 (t, *J* = 7.2 Hz, 2H), 1.91-1.74 (m, 3H), 1.56 (m, 2H), 0.93 (s, 9H), 0.93 (d, *J* = 6.5 Hz, 3H), 0.91 (d, *J* = 6.8 Hz, 3H), 0.11 (s, 3H), 0.07 (s, 3H).

^{13}C NMR: 100 MHz ($CDCl_3$) δ = 174.4, 173.4, 75.9, 72.5, 51.6, 44.8, 38.0, 31.2, 25.7, 24.9, 23.9, 23.4, 22.3, 17.9, -4.9, -5.2.

IR: (film, υ cm^{-1}) 3411, 2965, 2934, 1729, 1640, 1535.

HRMS: ESI calc. for $C_{16}H_{34}NO_4Si$ $[M+H]^+$:332.2251. Found 332.2252 $[M+H]^+$.



To a stirred solution of carboxylic acid (162 mg, 0.49 mmol, 1.0 eq.) in DMF (2.5 mL) at 23 °C was added *N*,*N*-diisopropylethylamine (0.36 mL, 2.0 mmol, 4.0 eq.) followed by HBTU (224 mg, 0.59 mmol, 1.2 eq.). After 5 min, *N*-acetylcysteamine (0.06 mL, 0.59 mmol, 1.2 equiv.) was added in one portion. After 12 h, the reaction was diluted with 1:1 hexanes:EtOAc and washed with sat. aqueous LiCl (3 x 50 mL). The organic layer was dried over sodium sulfate and concentrated *in vacuo*. The crude material was chromatographed on silica gel eluting with 1:2 hexanes:EtOAc to afford the title compound (128 mg, 60%) as a pale oil.

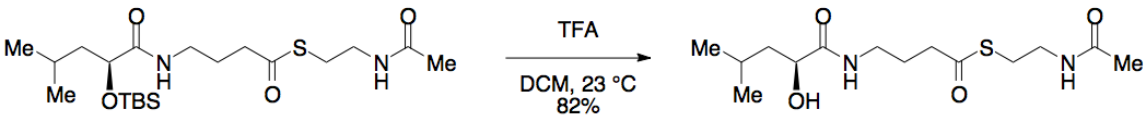
R_f = 0.19 (1:2 hexanes:EtOAc)

^{1}H NMR: 400 MHz ($CDCl_3$) δ = 6.63 (bs, 1H), 6.33 (bs, 1H) 4.13 (dd, *J* = 6.8 Hz, *J* = 4.8 Hz, 1H), 3.44 (q, *J* = 12.3 Hz, *J* = 6.2 Hz, 2H), 3.29 (qd, *J* = 6.8 Hz, *J* = 2.1 Hz, 2H) 3.04 (t, *J* = 6.2 Hz, *J* = 5.8 Hz, 2H), 2.61 (t, *J* = 7.2 Hz, 2H), 1.97 (s, 3H), 1.87 (p, *J* = 7.2 Hz, 2H), 1.78 (m, 1H), 1.56 (m, 1H), 0.94 (s, 9H), 0.92 (d, *J* = 6.5 Hz, 3H), 0.91 (d, *J* = 6.8 Hz, 3H), 0.11 (s, 3H), 0.08 (s, 3H).

^{13}C NMR: 100 MHz ($CDCl_3$) δ = 198.6, 174.4, 170.5, 72.4, 44.7, 40.6, 38.9, 37.2, 28.6, 25.6, 25.4, 23.9, 23.3, 22.9, 22.2, 17.8, -4.9, -5.3.

IR: (film, υ cm^{-1}) 3326, 2954, 1694, 1659, 1551, 1519.

HRMS: ESI calc. for $C_{20}H_{41}N_2O_4SSi$ $[M+H]^+$:433.2551. Found 433.2548 $[M+H]^+$.



To a stirred solution of TBS-protected thioester (153 mg, 0.37 mmol, 1.0 eq.) in DCM (3.7 mL) at 23 °C was added neat TFA (84 μL, 1.1 mmol, 3.0 eq.) in one portion. At 6 h, the reaction was concentrated *in vacuo*. The crude material was chromatographed on silica gel eluting with 3% MeOH in DCM to 4:1 DCM:MeOH to afford the title compound (93 mg, 82%) as a clear oil.

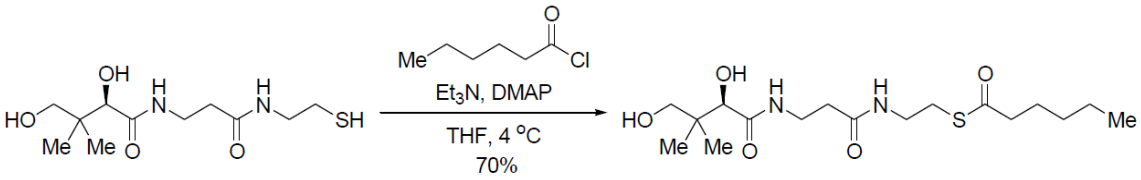
R_f = 0.25 (3% MeOH in DCM)

^{1}H NMR: 400 MHz ($CDCl_3$) δ = 6.67 (bs, 1H), 6.20 (bs, 1H), 4.14 (dd, *J* = 9.9 Hz, *J* = 3.1 Hz, 1H), 3.45 (q, *J* = 12 Hz, *J* = 5.8 Hz, 2H), 3.33 (qd, *J* = 6.8 Hz, *J* = 2.7 Hz, 2H), 3.03 (m, 2H), 2.63 (t, *J* = 6.8 Hz, 2H), 1.97 (s, 3H), 1.92 (dt, *J* = 6.8 Hz, *J* = 2.7 Hz, 1H), 1.84 (m, 1H), 1.66 (m, 2H), 1.53 (m, 2H), 0.97 (d, *J* = 6.5 Hz, 3H), 0.96 (d, *J* = 6.5 Hz, 3H).

^{13}C NMR: 100 MHz ($CDCl_3$) δ = 199.2, 175.2, 170.9, 70.6, 43.7, 40.8, 39.1, 37.7, 28.8, 25.2, 24.4, 23.5, 23.0, 21.3.

IR: (film, υ cm^{-1}) 3361, 2958, 1683, 1659, 1535.

HRMS: ESI calc. for $C_{14}H_{27}N_2O_4S$ $[M+H]^+$:319.1962. Found 341.1510 $[M+Na]^+$.



To a stirred solution of D-pantetheine (111 mg, 0.40 mmol, 1 eq.) dissolved in 8 mL THF and 2 mL Et_3N, DMAP was added (20 mg, 0.16 mmol, 0.4 eq.), and the solution was brought to 4 °C. Hexanoyl chloride (100 mg, 0.74 mmol, 1.9 eq.) was dissolved in 2 mL THF and added dropwise over 2 min to the stirring D-pantetheine solution. The solution was stirred 1 h at 4 °C and then 1 h at 22 °C. Solvent was removed *in vacuo*, and the remaining material was resuspended in acetone and filtered. The acetone was removed *in vacuo*, and the clear oil was chromatographed on silica gel eluting with 1:1 EtOAc:acetone to afford the title compound (105 mg, 70%) as a clear oil.

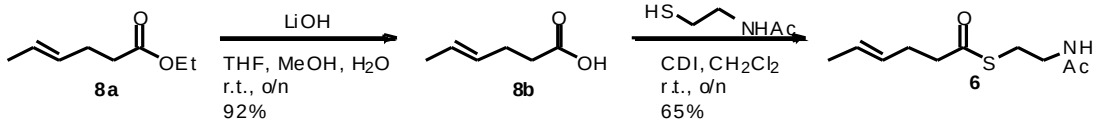
R_f = 0.3 (1:1 EtOAc:acetone)

^{1}H NMR: 400 MHz ($CDCl_3$) δ = 7.35 (t, *J* = 5.3 Hz, 1H), 6.18 (t, *J* = 5.3 Hz, 1H), 3.99 (s, 1H), 3.59-3.32 (m, 6H), 3.08-2.95 (m, 2H), 2.56 (t, *J* = 8.0 Hz, 2H), 2.41 (t, *J* = 5.8 Hz, 2H), 1.65 (p, *J* = 7.6 Hz, 2H), 1.34-1.23 (m, 4H), 1.02 (s, 3H), 0.92 (s, 3H), 0.89 (t, *J* = 7.2 Hz, 3H).

HRMS: ESI calc. for $C_{17}H_{33}N_2O_5S$ $[M+H]^+$:376.2110. Found 399.1924 $[M+Na]^+$.

Synthesis of ligands for mass spectrometry analysis of KS acylation

All chemicals were purchased from commercial sources and were used without further purification. Solvents were dried as described (Hünig *et al.*, 2006). All reactions were carried out under an atmosphere of argon unless otherwise noted. ^{1}H NMR and ^{13}C NMR were recorded on a Bruker dpx 400 (400 MHz/100 MHz) or Bruker dpx 300 (300 MHz/75 MHz) spectrometer at room temperature. Chemical shifts were reported in parts per million (ppm) and the residual solvent peak was used as internal reference: ^{1}H-NMR ($CHCl_3$: 7.26 ppm), ^{13}C-NMR ($CDCl_3$: 77.16 ppm). High resolution mass spectra (HRMS) were recorded with a Bruker micrOTOF-Q mass spectrometer. IR was recorded on a Thermo Smart Orbit Nicolet 380 FT-IR spectrometer. TLC was performed on silica gel plates (Merck Silica gel 60 F_{254}) and visualized with Seebach reagent (12.5 g phosphomolybdic acid, 5.0 g cerium(IV) sulphate tetrahydrate, 16.0 mL water, 450.0 mL concentrated sulfuric acid). Flash column chromatography was performed on silica gel (Merck-Kieselgel 60, 40–60 mesh). Abbreviations used: CDI, carbonyldiimidazole; EtOAc, ethyl acetate; THF, tetrahydrofuran.



8a **8b** **6**

(*E*)-Hex-4-enoic acid (8b) (Smith et al., 1997)

LiOH (4.26 g, 177.87 mmol, 25 eq.) was added to a stirred solution of (*E*)-ethyl hex-4-enoate (Noack & Göttlich, 2002) (**8a**) (1.00 g, 7.03 mmol) in THF/MeOH/H_2O (290 mL, 2:2:1) at r.t., and the mixture was stirred at r.t. overnight. The organic solvents were removed under reduced pressure, and the aqueous phase was acidified with aqueous 0.1 M $NaHSO_4$ to pH = 2. The aqueous phase was extracted with EtOAc (3 x 50 mL), and the combined organic phases were washed with brine (20 mL). The organic phase was dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The acid **8b** (0.74 g, 92%) was obtained as a colorless oil and used without further purification.

^{1}H NMR: (400 MHz, $CDCl_3$) (Jefford & Wang, 1987) δ 1.56 – 1.60 (m, 3H), 2.20 - 2.28 (m, 2H), 2.31 – 2.37 (m, 2H), 5.29 – 5.50 (m, 2H).

(*E*)-*S*-(2-Acetamidoethyl) hex-4-enethioate (6) (Deska et al., 2011)

CDI (1.74 g, 10.74 mmol, 1.2 eq.) was added to a stirred solution of (*E*)-hex-4-enoic acid **8b** (1.02 g, 8.95 mmol) in anhydrous CH_2Cl_2 (30 mL) at room temperature. After 10 min, *N*-acetylcysteamine (1.14 mL, 1.28 g, 10.74 mmol, 1.2 eq.) was added, and the reaction mixture was stirred at room temperature overnight. The solvent was removed under reduced pressure, and the residue was dissolved in EtOAc (20 mL). The organic phase

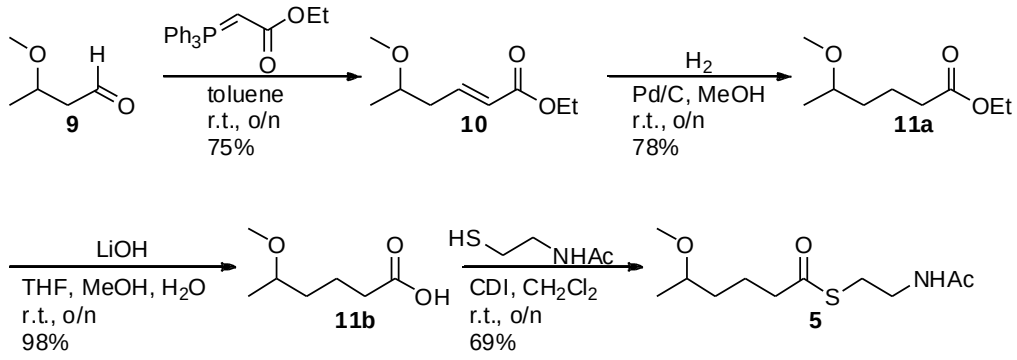
was washed with water (3 x 5 mL) and brine (2 x 5 mL). The organic phase was dried (Na_2SO_4), and the solvent was removed under reduced pressure. Purification by column chromatography (SiO_2, EtOAc, R_f = 0.43) gave thioester **6** (1.26 g, 65%) as a white, waxy solid.

^{1}H NMR: (400 MHz, $CDCl_3$) δ 1.61 (dd, J = 6.3, 1.4 Hz, 3H), 1.94 (s, 3H), 2.27 – 2.35 (m, 2H), 2.60 (t, J = 6.9 Hz, 2H), 3.00 (t, J = 6.4 Hz, 2H), 3.36 - 3.44 (m, 2H), 5.30 - 5.40 (m, 1H), 5.41 - 5.52 (m, 1H), 6.03 (brs, 1H)

^{13}C NMR: (75 MHz, $CDCl_3$) δ 18.0 (CH_3), 23.3 (CH_3), 28.53 (CH_2), 28.54 (CH_2) 39.8 (CH_2), 44.0 (CH_2), 126.8 (CH), 128.5 (CH), 170.4 (C), 199.6 (C);

IR: (neat) ν_{max} 598, 963, 1036, 1287, 1372, 1435, 1545, 1650, 1688, 2933, 3272 cm^{-1};

HRMS: *m/z* 238.0877 (calcd. for $C_{10}H_{17}NO_2SNa$ 238.0872).



(*E*)-Ethyl 5-methoxyhex-2-enoate (10) (Brückner et al., 2008)

Ethyl 2-(triphenylphosphoranylidene)acetate (1.28 g, 3.67 mmol) was added to a solution of 3-methoxybutanal (Münster *et al.*, 2012) (**9**) (0.34 g, 3.67 mmol) in anhydrous toluene (12 mL) at r.t., and the mixture was stirred at r.t. overnight. The solvent was removed

under reduced pressure, and the residue was suspended in *c*-hexane (50 mL). The suspension was filtered through a plug of Celite® 545, and the filtrate was concentrated under reduced pressure. Purification by column chromatography (SiO_2, EtOAc/*c*-hexane, 1/3, R_f = 0.55) gave the ester **10** (0.45 g, 75%) as a colorless oil.

^{1}H NMR: (400 MHz, $CDCl_3$) δ .15 (d, J = 6.2 Hz, 3H), 1.27 (t, J = 7.1 Hz, 3H), 2.22 – 2.52 (m, 2H), 3.32 (s, 3H), 3.38 – 3.50 (m, 1H), 4.17 (q, ^3J= 7.1 Hz, 2H), 5.85 (dt, J = 15.7, 1.5 Hz, 1H), 6.93 (dt, J = 15.7, 7.3 Hz, 1H)

^{13}C NMR: (75 MHz, $CDCl_3$) δ 14.4 (CH_3), 19.2 (CH_3), 39.0 (CH_2), 56.3 (CH_3), 60.3 (CH_2), 75.7 (CH), 123.5 (CH), 145.3 (CH), 166.5 (C)

IR: (neat) ν_{max} 1040, 1090, 1172, 1264, 1716, 2357, 2977 cm^{-1}

HRMS: *m/z* 195.0990 (calcd. for $C_9H_{16}O_3Na$ 195.0992).

Ethyl 5-methoxyhexanoate (11a)

A suspension of the unsaturated (*E*)-ethyl 5-methoxyhex-2-enoate (**10**) (0.84 g, 4.94 mmol) and Pd/C (0.09 g, 10 mol%) in anhydrous methanol (8 mL) was stirred under hydrogen atmosphere (1 bar) at room temperature for 16 h. The suspension was filtered through a plug of Celite®545 and rinsed with EtOAc (20 mL). The solvents of the combined organic fractions were removed under reduced pressure, yielding the desired ester **11a** (0.66 g, 78%) as a colorless oil.

^{1}H NMR: (400 MHz, $CDCl_3$) δ 1.12 (d, J = 6.1 Hz, 3H), 1.24 (t, J = 7.1 Hz, 3H), 1.37 – 1.58 (m, 2H), 1.58 – 1.78 (m, 2H), 2.30 (t, J = 7.4 Hz, 2H), 3.14 – 3.41 (m, 4H), 4.12 (q, J = 7.1 Hz, 2H)

^{13}C NMR (75 MHz, $CDCl_3$) δ 14.4 (CH_3), 19.1 (CH_3), 21.1 (CH_2), 35.5 (CH_2), 35.8 (CH_2), 56.1 (CH_3), 60.4 (CH_2), 76.5 (CH), 173.8 (C)

IR: (neat) ν_{max} 1082, 1164, 1372, 1732, 2359, 2970 cm^{-1}

HRMS: *m/z* 197.1146 (calcd. for $C_9H_{18}O_3Na$ 197.1148).

5-Methoxyhexanoic acid (11b)

Acid **11b** (0.57 g, 98%) was obtained from ethyl 5-methoxyhexanoate (**11a**) (0.67 g, 3.86 mmol), according the procedure described for acid **8b**, as a colorless oil

^{1}H NMR: (400 MHz, $CDCl_3$) δ 1.14 (d, J = 6.1 Hz, 3H), 1.29 – 1.83 (m, 4H), 2.37 (t, J = 7.3 Hz, 2H), 3.25 – 3.48 (m, 4H), 10.59 (br s, 1H, H-8)

^{13}C NMR: (75 MHz, $CDCl_3$) δ 19.0 (CH_3), 20.8 (CH_2), 34.1 (CH_2), 35.7 (CH_2), 56.1 (CH_3), 76.6 (CH), 177.6 (C)

IR: (neat) ν_{max} 1078, 1135, 1169, 1706, 2970 cm^{-1}

HRMS: *m/z* 145.0831 (calcd. for $C_7H_{14}O_3$ 145.0835).

***S*-(2-acetamidoethyl) 5-methoxyhexanethioate (5)**

Thioester **5** was prepared from 5-methoxyhexanoic acid (**11b**) (0.57 g, 3.89 mmol) according to the procedure described for thioester **6**. Purification by column

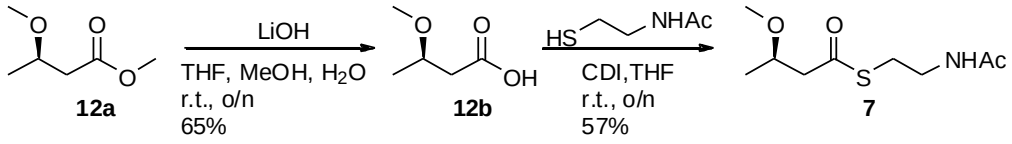
chromatography (SiO_2, EtOAc, R_f = 0.33) gave the thioester **5** (0.66 g, 69%) as a white, waxy solid.

^{1}H NMR: (400 MHz, $CDCl_3$) δ 1.11 (d, J = 6.1 Hz, 3H), 1.29 – 1.58 (m, 2H), 1.59 – 1.88 (m, 2H), 1.95 (s, 3H), 2.58 (t, J = 7.4 Hz, 2H), 3.01 (t, J = 6.4 Hz, 2H), 3.19 – 3.35 (m, 4H), 3.37 – 3.46 (m, 2H), 5.91 (br s, 1H)

^{13}C NMR: (75 MHz, $CDCl_3$) δ 19.0 (CH_3), 21.8 (CH_3), 23.3 (CH_2), 28.6 (CH_2), 35.6 (CH_2), 39.8 (CH_2), 56.1 (CH_3), 76.4 (CH), 170.4 (C), 200.1 (C)

IR: (neat) ν_{max} 598, 1087, 1135, 1286, 1372, 1543, 1651, 1686, 2930, 3290 cm^{-1}

HRMS: *m/z* 270.1142 (calcd. for $C_{11}H_{21}O_3SNa$ 270.1140).



(*R*)-3-methoxybutanoic acid (12b)

Acid **12b** was prepared from (*R*)-methyl 3-methoxybutanoate (Page *et al.*, 2004) (**12a**) (0.94 g, 7.11 mmol) according the procedure described for acid **8b**. The acid **12b** (0.55 g, 65%) was obtained as a colorless oil and used without further purification.

^{1}H NMR: (400 MHz, $CDCl_3$) (Zhu & Burgess, 2008) δ 1.23 (d, J = 6.0 Hz, 3H), 2.45 (dd, J = 15.4, 5.4 Hz, 1H), 2.59 (dd, J = 15.4, 7.2 Hz, 1H), 3.35 (s, 3H), 3.78 (ddd, J = 7.2, 6.0 Hz, 5.4 Hz, 1H), 9.99 (br s, 1H)

^{13}C NMR: (75 MHz, $CDCl_3$) δ 19.1 (CH_3), 41.6 (CH_2), 56.5 (CH_3), 73.6 (CH), 176.9 (C-5).

(*R*)-*S*-(2-acetamidoethyl) 3-methoxybutanethioate (7)

Thioester **7** was prepared from (*R*)-3-methoxybutanoic acid (**12b**) (0.55 g, 4.67 mmol) according to the procedure described for thioester **6**. Anhydrous THF (10 mL) was used instead of CH_2Cl_2. Purification by column chromatography (SiO_2, EtOAc, R_f = 0.34) gave thioester **7** (0.58 g, 57%) as a white, waxy solid.

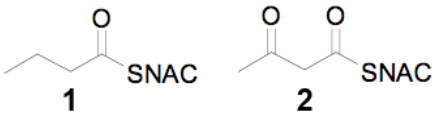
^{1}H NMR: (400 MHz, $CDCl_3$) δ 1.19 (d, J = 6.1 Hz, 3H), 1.93 (s, 3H), 2.58 (dd, J = 15.1, 5.7 Hz, 1H), 2.81 (dd, J = 15.1, 7.2 Hz, 1H), 3.01 (t, J = 6.5 Hz, 2H), 3.29 (s, 3H), 3.37 – 3.44 (m, 2H), 3.77 (ddd, J = 7.2, 6.1, 5.7 Hz, 1H), 6.05 (br s, 1H)

^{13}C NMR: (75 MHz, $CDCl_3$) δ 19.2 (CH_3), 23.2 (CH_3), 28.7 (CH_2), 39.6 (CH_2), 50.8 (CH_2), 56.5 (CH_3), 73.8 (CH), 170.4 (C), 197.8 (C)

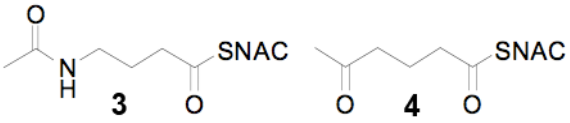
IR: (neat) ν_{max} 997, 1086, 1284, 1373, 1544, 1651, 2360, 2931, 3283 cm^{-1}

HRMS: *m/z* 220.0996 (calcd. for $C_9H_{18}NO_3S$ 220.1007).

Substrates **1** and **2** were prepared following previously established routes (Jenner *et al.*, 2013).



Substrates **3** and **4** were prepared following previously established routes (Kohlhass *et al.*, 2013).



ACKNOWLEDGEMENTS

Instrumentation and technical assistance for crystallographic work were provided by Dr. Art Monzingo and the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President

and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank the National Institutes of Health (GM106112) and the Welch Foundation (F-1712) for supporting this research (A.T.K.). We are grateful for financial support from the Biotechnology and Biological Science Research Council (BBSRC) for studentship to M.J., the Leverhulme Trust for funding (code RPG-2012-578) to J.A. and N.J.O., and the DFG (SFB 642) to J.P. We thank Sarah Frank for synthesizing short-chain NAC derivatives, and Maria Person at the University of Texas at Austin Proteomics Facility for protein mass spectrometry and analysis.

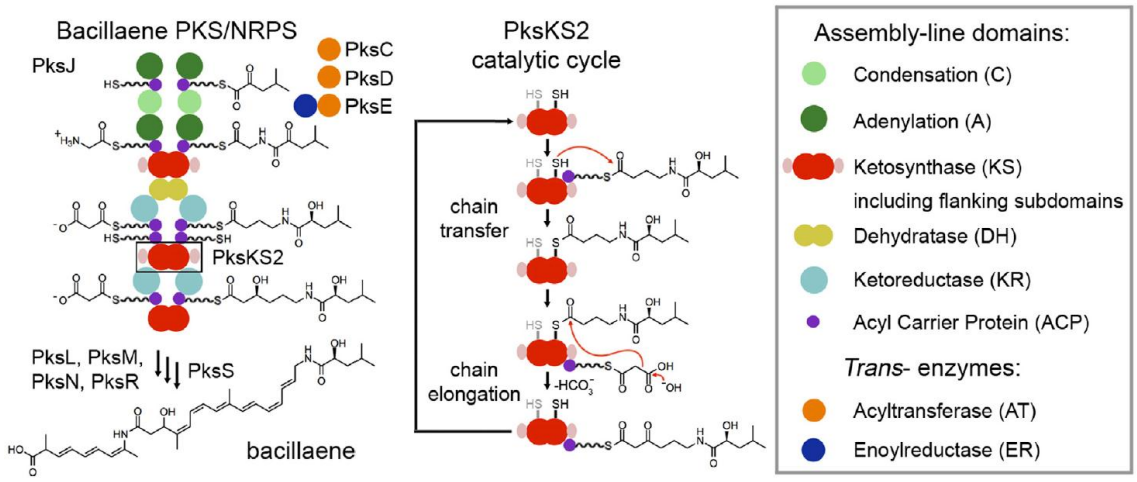


Figure 6-1. A bacillaene synthase ketosynthase. PksKS2 is in PksJ, the first subunit of a hybrid *trans*-AT polyketide synthase/nonribosomal peptide synthetase (PKS/NRPS) from *Bacillus subtilis*. Discrete polypeptides containing ATs and an ER (PksC, PksD, and PksE) cooperate *in trans* with the assembly line polypeptides (PksJ, PksL, PksM, PksN, and PksR) in the construction of bacillaene. In the catalytic cycle of PksKS2, an intermediate is transferred from an ACP of the previous module before the malonyl-bound ACP of the same module is condensed with it. See also Figure 6-2.

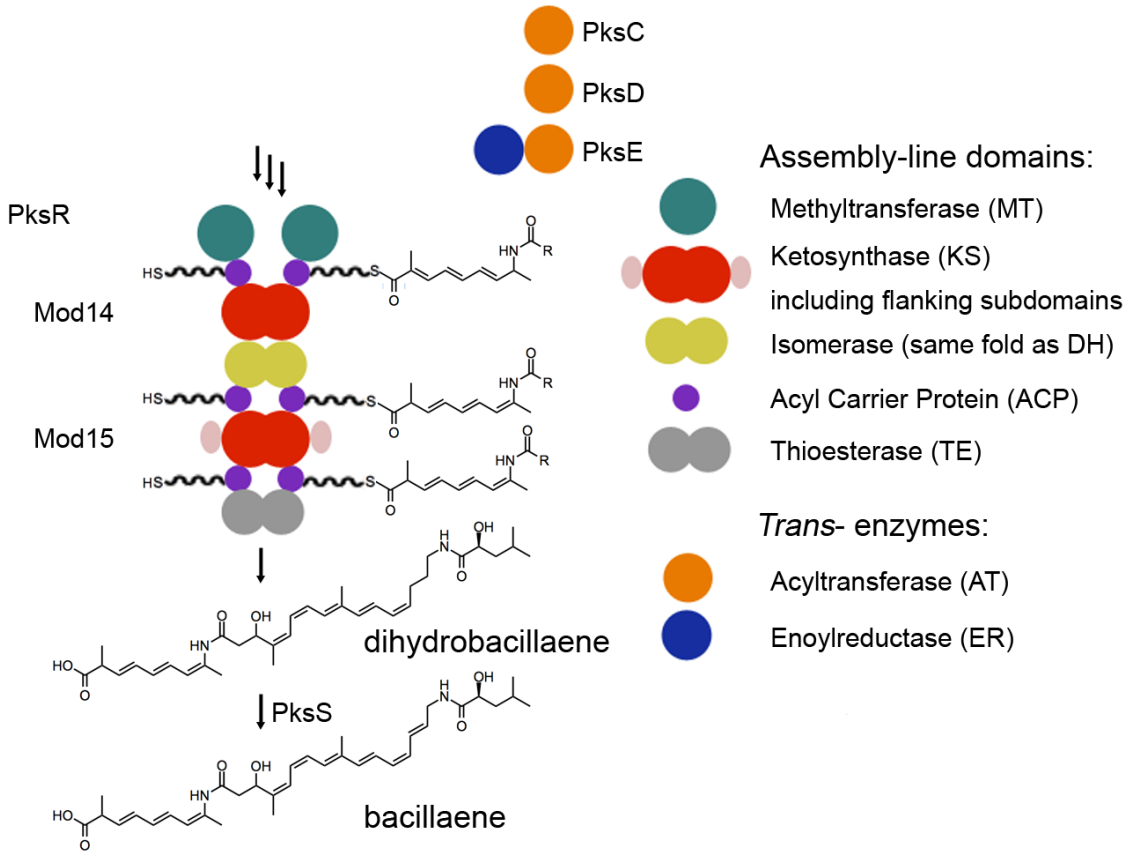


Figure 6-2. KS gatekeeping. The last KS of bacillaene synthase is condensation-incompetent and apparently plays a gatekeeping role to select for the intermediate that has been productively isomerized by the previous module.

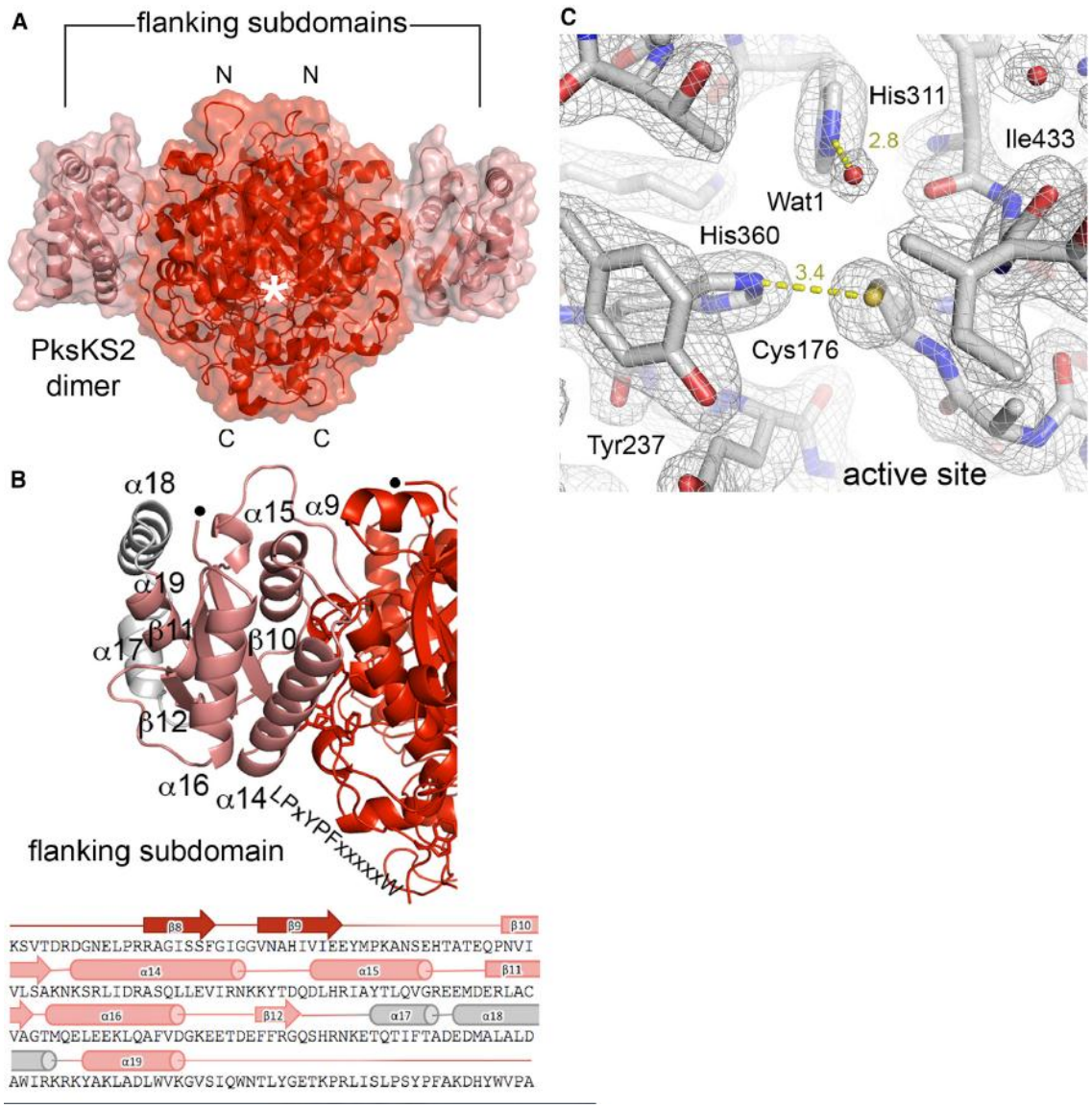
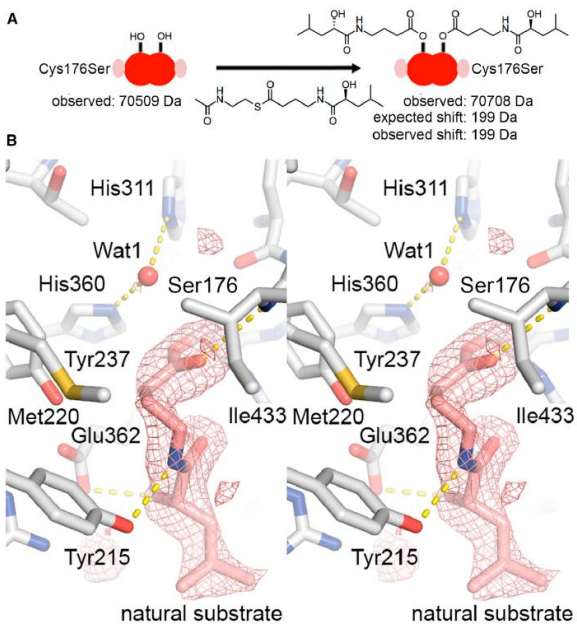


Figure 6-3. PksKS2 architecture. **(a)** Like ~90% of *trans*-AT KSs, PksKS2 possesses a flanking subdomain equivalent to the KS/AT adapter in *cis*-AT PKSs. One active site is indicated with an asterisk. **(b)** The flanking subdomain contains two loosely-connected helices, α17 and α18, in the position that AT is inserted in *cis*-AT PKSs. The interaction between the body of the KS and the flanking subdomain is primarily mediated by the LPxYPF motif; nine unstructured residues link the body of the KS to the flanking subdomain as indicated by the circles. **(c)** The 1.95-Å resolution $2F_o$-F_c electron density maps, contoured at 1.5 r.m.s.d., show the highest-resolution view yet available for a KS from a type I PKS. Transfer of the polyketide intermediate to Cys176 is aided by His360. That condensation-incompetent KSs usually lack the histidine equivalent to His311 implies its role in the chain elongation reaction. See also Figure 6-4.

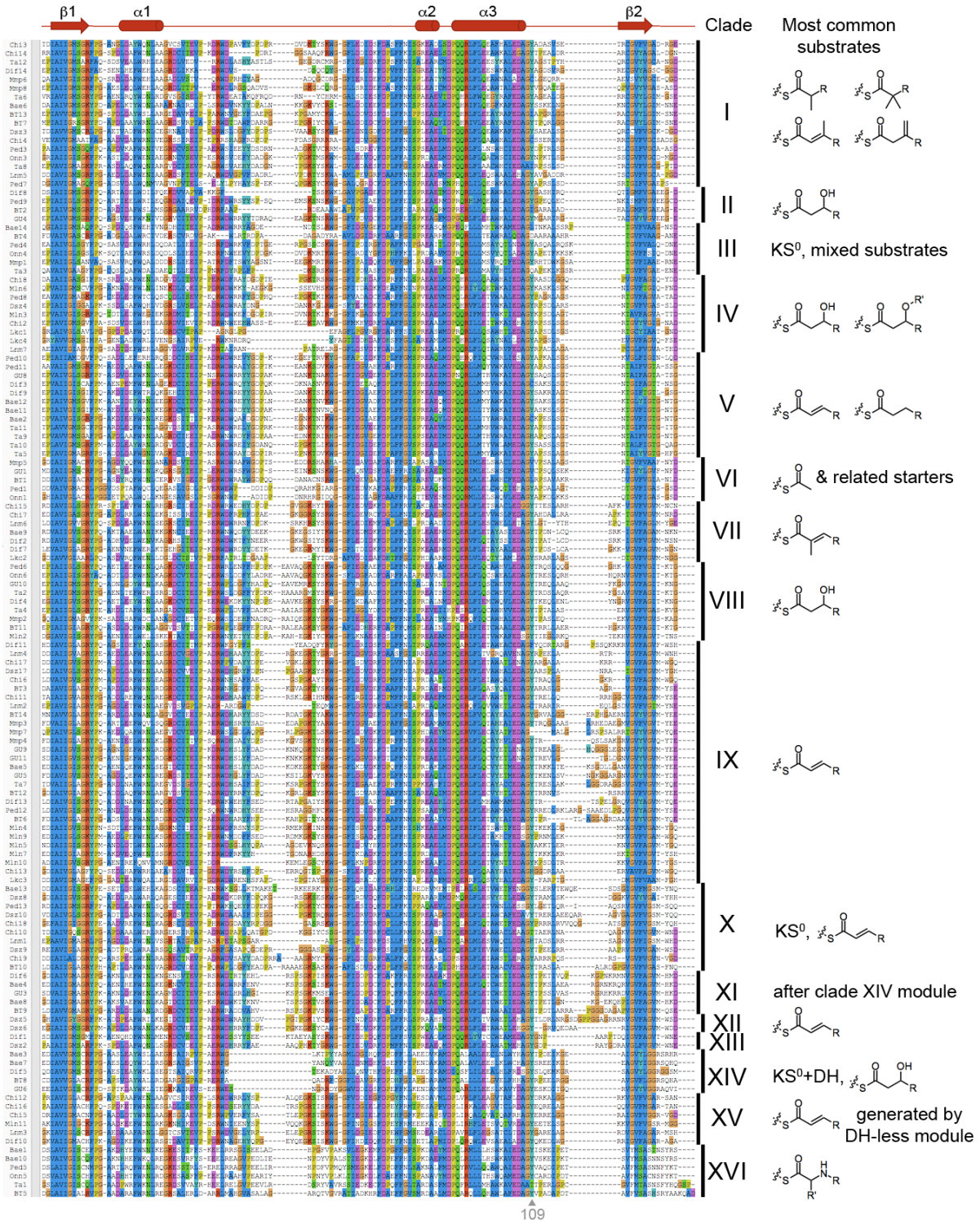
Figure 6-4. Secondary structure assignment. The secondary structure features of PksKS2 were assigned using the Dictionary of Secondary Structure of Proteins (Kabsch & Sander, 1983).



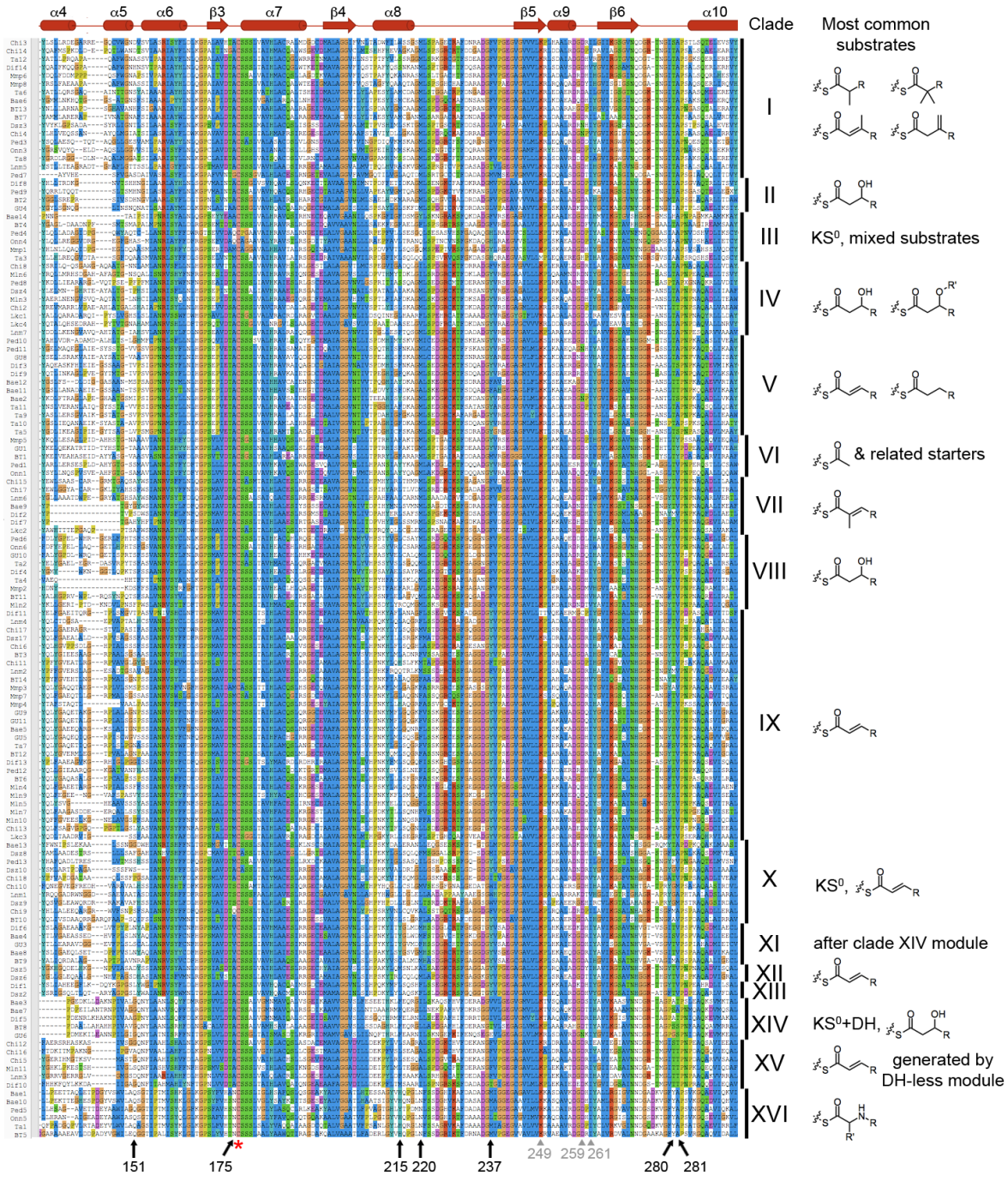
C

Clade	151'	175	215	220	237	280	281	317	321	362	363	365	433	434
I	A	A	M	M	F	T	A	L	I	A	A/T	A	F	S
II	H	A	M	M	L	T	A	L	V/I	F	A	S	M	S
III	A	A/Q/M	L	Q	H	A	A	V	A	E	S	S	A/F	G
IV	S	A	A/L/F	M	Y	T	A	L	I	E	P	A	F	G
V	V	A	F	M	Y	T	A	L	I	E	L	A	F	G
VI	A	A	F	M	Y	T	Y	K	I	E	A	A	F	G
VII	A	A	L	M	F	T	V	L	I	E	S	A	A	G
VIII	S	M	L	M	F	T	V	L	I	E	A	A	A	G
IX	S	M	L	F	Y	T	V	L	I	E	S	A	A	G
X	S	A	L	L	Y/W	G	A/V	L	I	E	A	S	A	G
XI	N	V	Y	M	Y	T	V	L	I	E	S	A	A	S/T
XII	D	A	L/M	A/M	Y	T	V	L	I	E	G/S	A	A	G
XIII	L	A	L	F	Y	T	V	L	V/I	E	A	A	A	G
XIV	G	A	F	L	V	A/S	S/T	V	L	L	C	E	D	G
XV	G	A	L	A	F	T	T	I	I	L	S	A	F	G
XVI	Q	N	H	N	M	Y	A	L	I	D	T	A	L	G

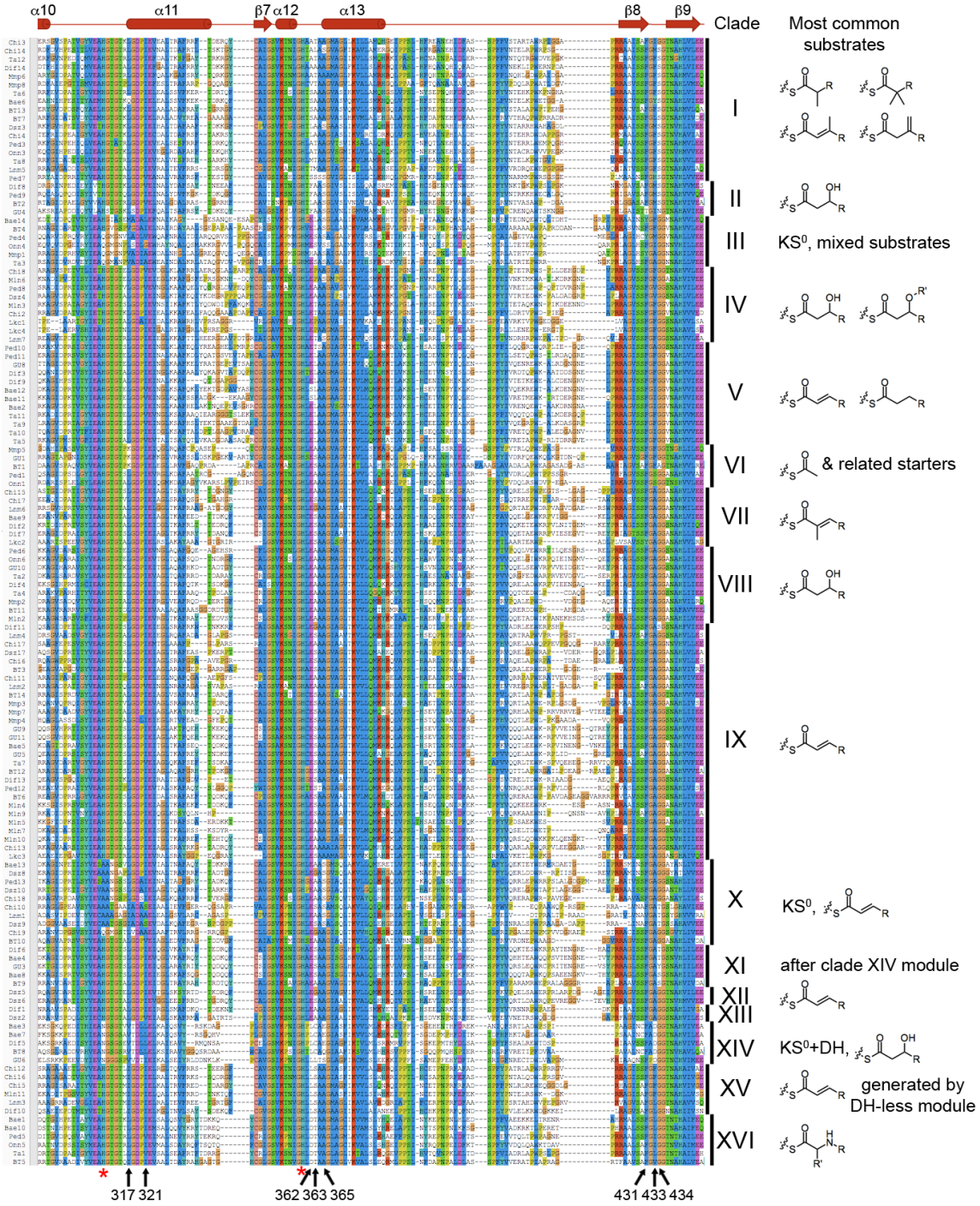
Figure 6-5. PksKS2 bound to substrates. **(a)** To crystallographically observe the natural substrate covalently bound at high occupancy, mutation of the catalytic cysteine to serine was necessary (crystals were soaked with the N-acetylcysteamine thioester of the natural substrate). The expected mass shift for the acyl group was observed by mass spectrometry. **(b)** A stereodiagram shows the F_o-F_c omit map (contoured at 3.0 r.m.s.d.; substrate atoms, represented in salmon, were omitted prior to map calculation) for the natural substrate of PksKS2. The ester carbonyl forms a hydrogen bond with Ile433. The side chains of Met220, Glu362, Ile433, and Tyr237 are nearest to the α- and β-carbons of the polyketide and may help confer specificity. Hydrogen bonds, indicated with yellow dashes, are each 2.9–3.1 Å. (**c**) Residues within 15 Å of the PksKS2 reactive cysteine that vary between the sixteen KS clades were identified. The most highly represented residues are shown for each clade. See also Figure 6-6 and Table 6-2.



(Figure 6-6 continued on next page.)



(Figure 6-6 continued on next page.)



(Figure 6-6 continued on next page.)

Figure 6-6. KS alignment. The KSs described in "Biosynthesis of polyketides by *trans*-AT polyketide synthases" (Piel, 2010) were aligned, and sequences were taken from the supplementary information found within "Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection" (Nguyen *et al.*, 2008). The secondary structure is from the crystal structure of PksKS2. Catalytic residues are indicated with asterisks. Residues consistently different from *cis*-AT PKSs are labeled grey. Residues within 15 Å of the catalytic cysteine and that vary between clades are labeled black. KS clades and their associated substrates are indicated on the right.

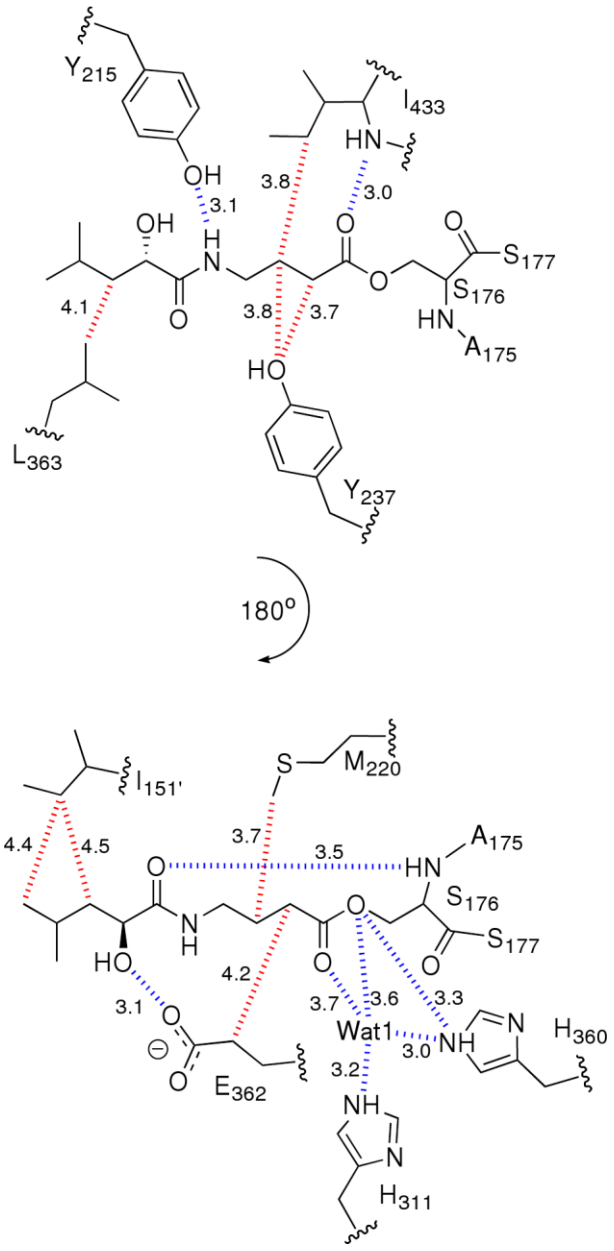


Figure 6-7. Interactions between PksKS2 active site residues and its natural substrate. Shown in cartoon form are the interactions made between the natural substrate for PksKS2 and each active site residue. Due to the large number of interactions, two angles are presented. Red dashes indicate steric interactions and blue dashes indicate hydrogen bonds, with distance units shown in Å.

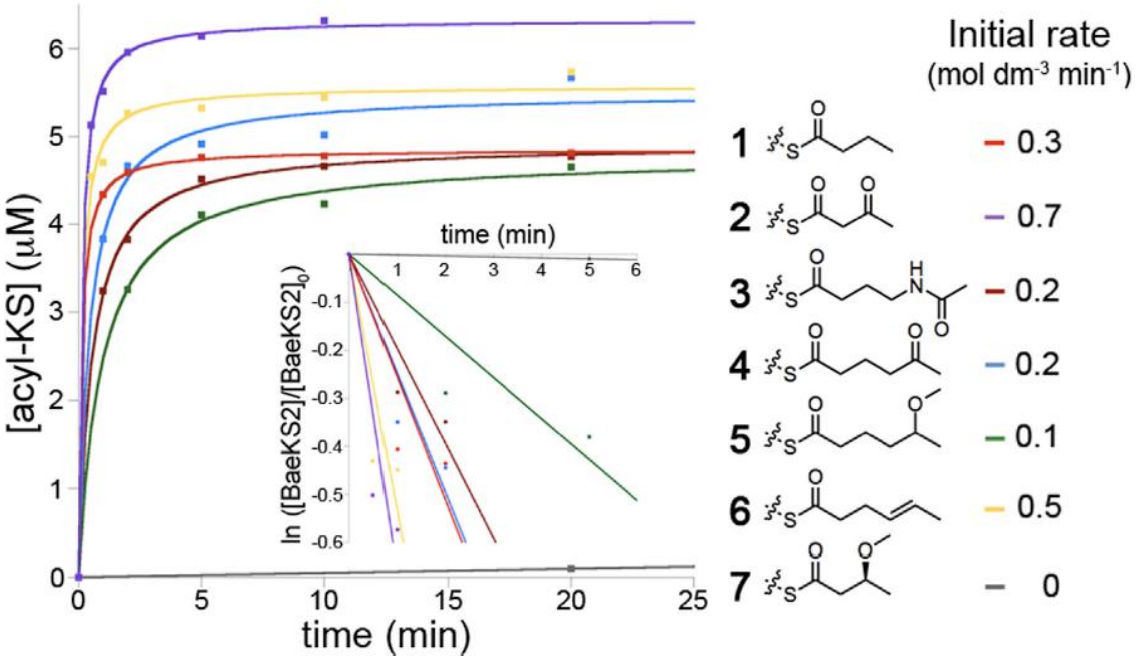
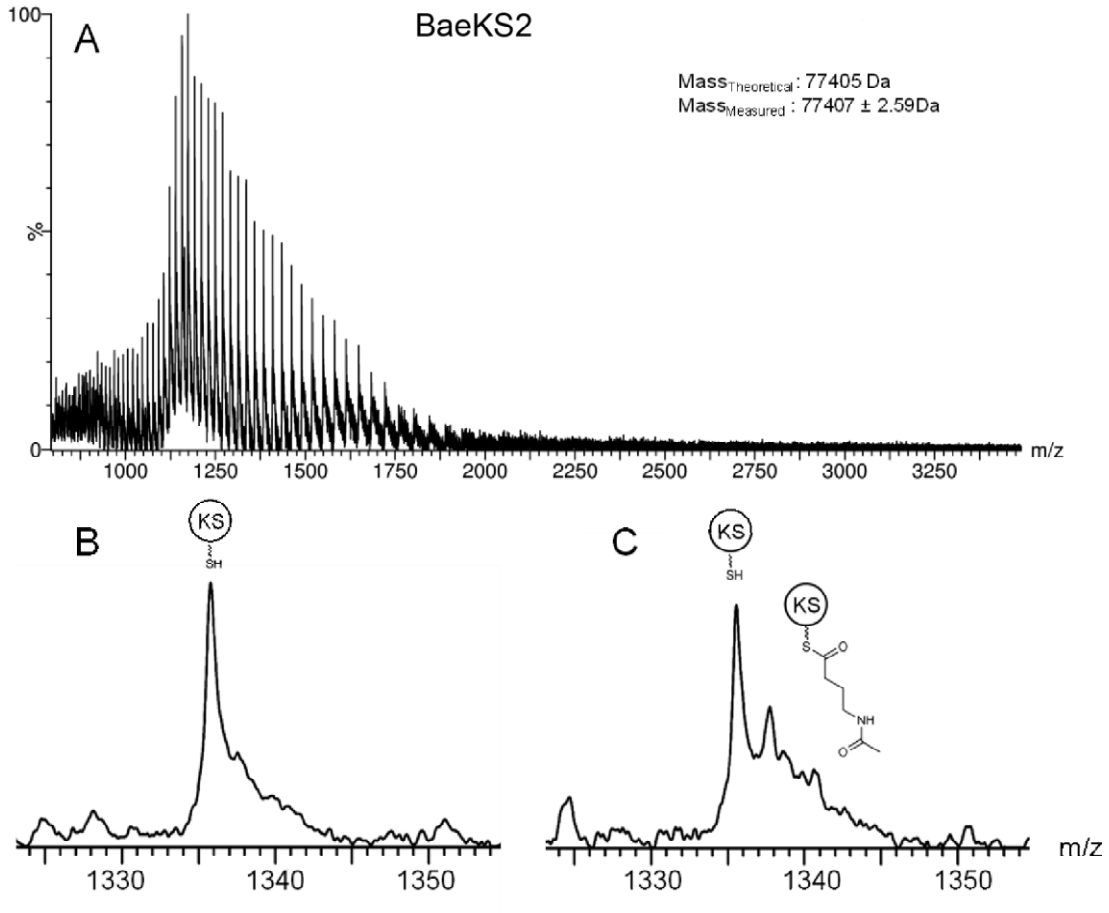
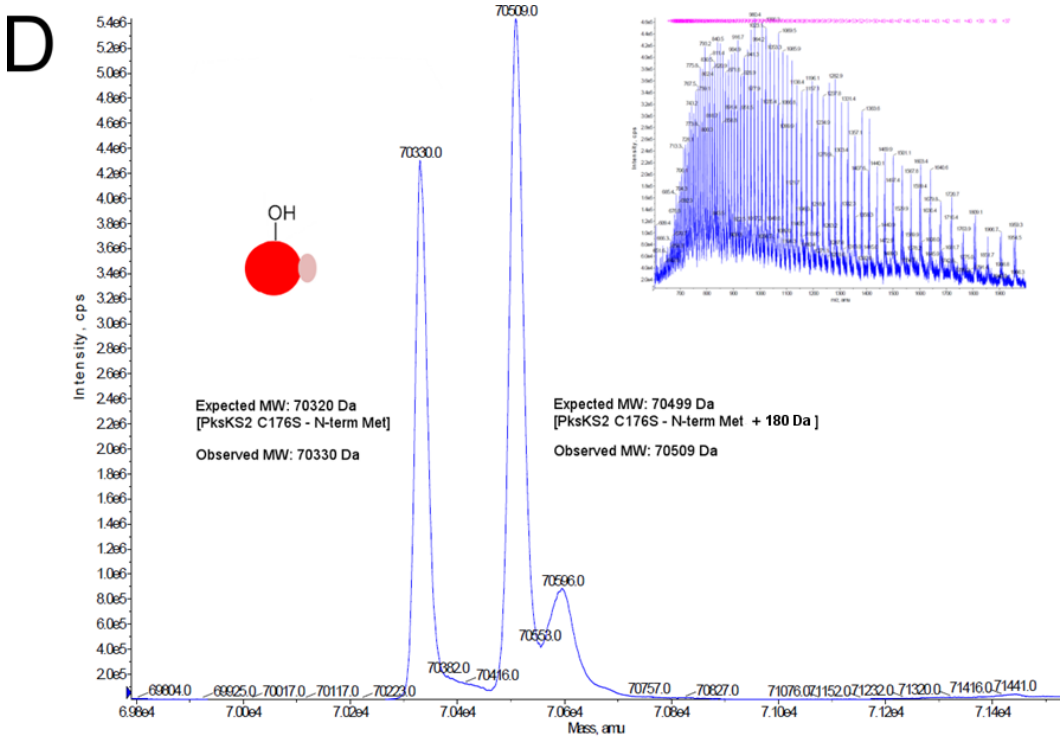
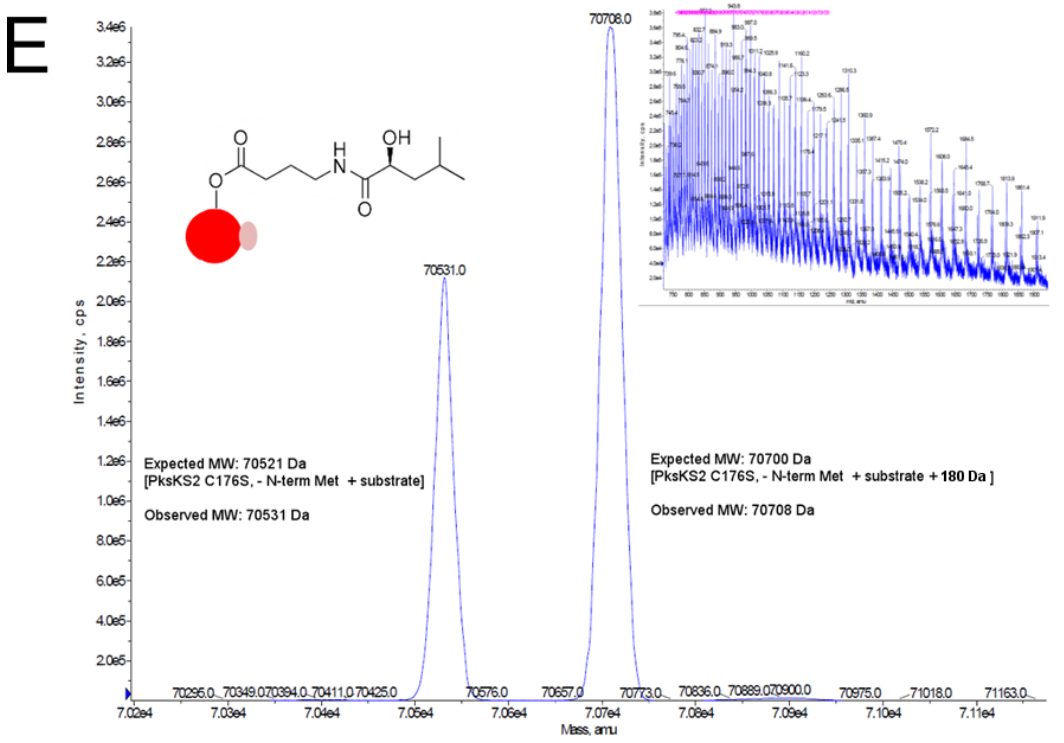


Figure 6-8. Examining KS specificity by mass spectrometry. Acyl-*S*-NACs **1-7** were incubated with BaeKS2. The formation of the acyl-enzyme intermediate was monitored through mass spectrometry. Early time points are reported in the subpanel. The specificity of BaeKS2 is quite broad; however, it does not tolerate a methoxy group at the β-position. See also Figure 6-9.



(Figure 6-9 continued on next page.)





(Figure 6-9 continued on next page.)

Figure 6-9. KS acylation measured by mass spectrometry. **(a)** Nano-ESI mass spectrum of BaeKS2 sprayed from 80:20 MeCN:H_2O 0.1% TFA, resulting in a measured mass of 77,407 Da. **(b)** The 58^+ charge state of BaeKS2 not incubated with substrate. **(c)** The 58^+ charge state of BaeKS2 following incubation with 2 mM thioester 4, clearly showing a mass shift corresponding to acylated BaeKS2. **(d)** ESI-mass spectrum (Figure insert) and mass reconstruction of Cys176Ser PksKS2. The mass observed corresponds to an N-terminal methionine cleavage, in addition to an anomalous glycosylation (+ 180 Da). No sugar moieties were visible in the high-resolution crystal structure, indicating that the site of glycosylation was either highly disordered, or was excluded from the crystals due to unfavorable crystal packing. **(e)** ESI-mass spectrum (Figure insert) and mass reconstruction of Cys176Ser PksKS2 after a 1 hour incubation with 10 mM of the natural substrate thioester, showing >99% active site ester formation.

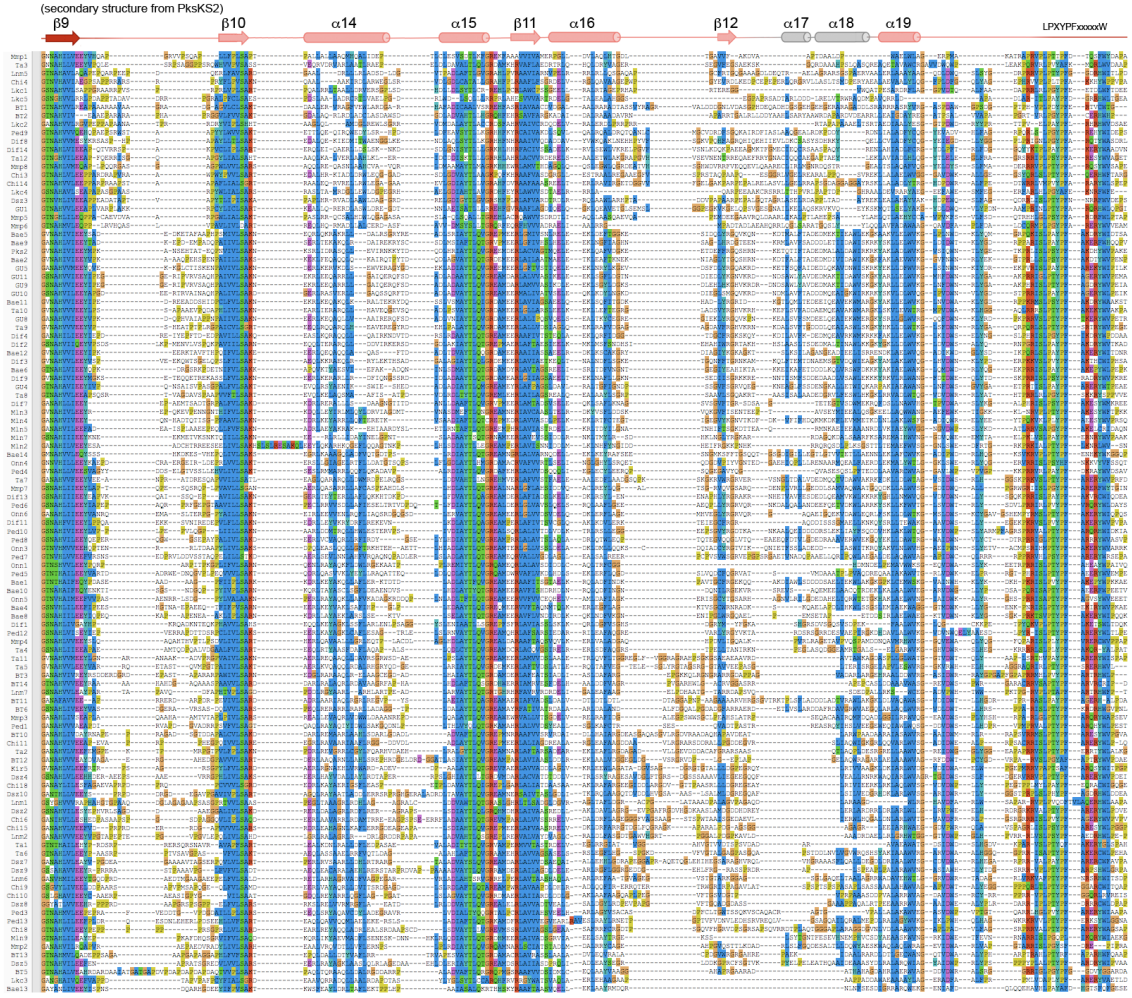


Figure 6-10. Flanking subdomain alignment. The flanking subdomains from KSs described in "Biosynthesis of polyketides by *trans*-AT polyketide synthases" (Piel, 2010) were aligned, and sequences were taken from the supplementary information found within "Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection" (Nguyen *et al.*, 2008).

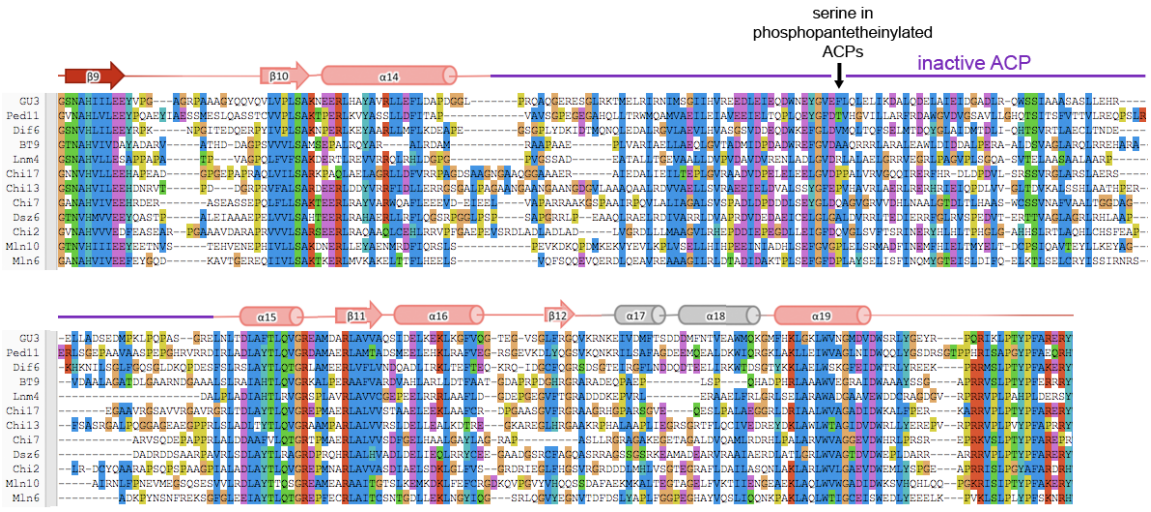
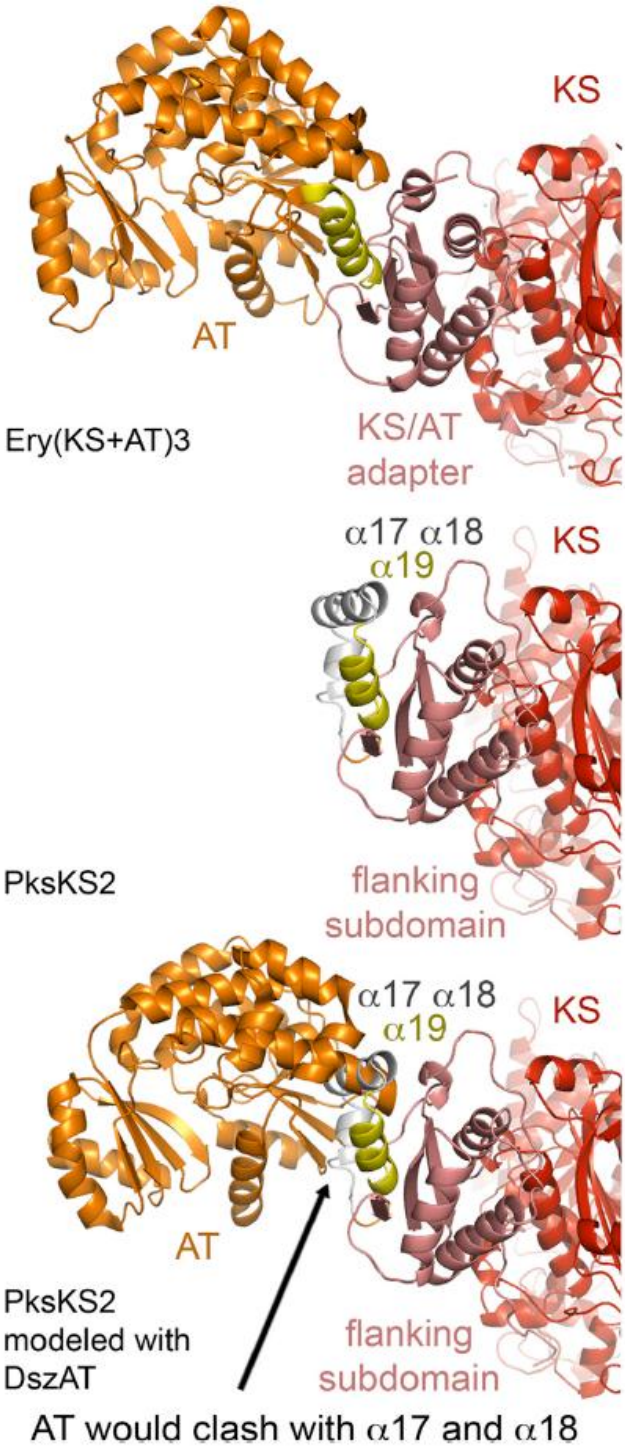


Figure 6-11. Flanking subdomains containing an inactive ACP. An alignment shows the insertion of inactive ACP domains between α14 and α15 of the flanking subdomain.



(Figure 6-12 continued on next page.)

Figure 6-12. Clashes from AT docking. If a discrete AT, such as the disorazole AT (DszAT, PDB Code: 3RGI), were to dock to PksKS2 in an equivalent position to what has been observed for *cis*-AT PKSs, such as Ery(KS+AT)3 (PDB Code: 2QO3) and the metazoan FAS, the AT would clash with α17 and α18 of the flanking subdomain; however, these loosely-connected helices could shift. Sequence alignments of several trans-AT domains reveal the absence of a helix equivalent to the final helix of Streptomyces coelicolor FabD, which could enable an interaction with α19 of the flanking subdomain.

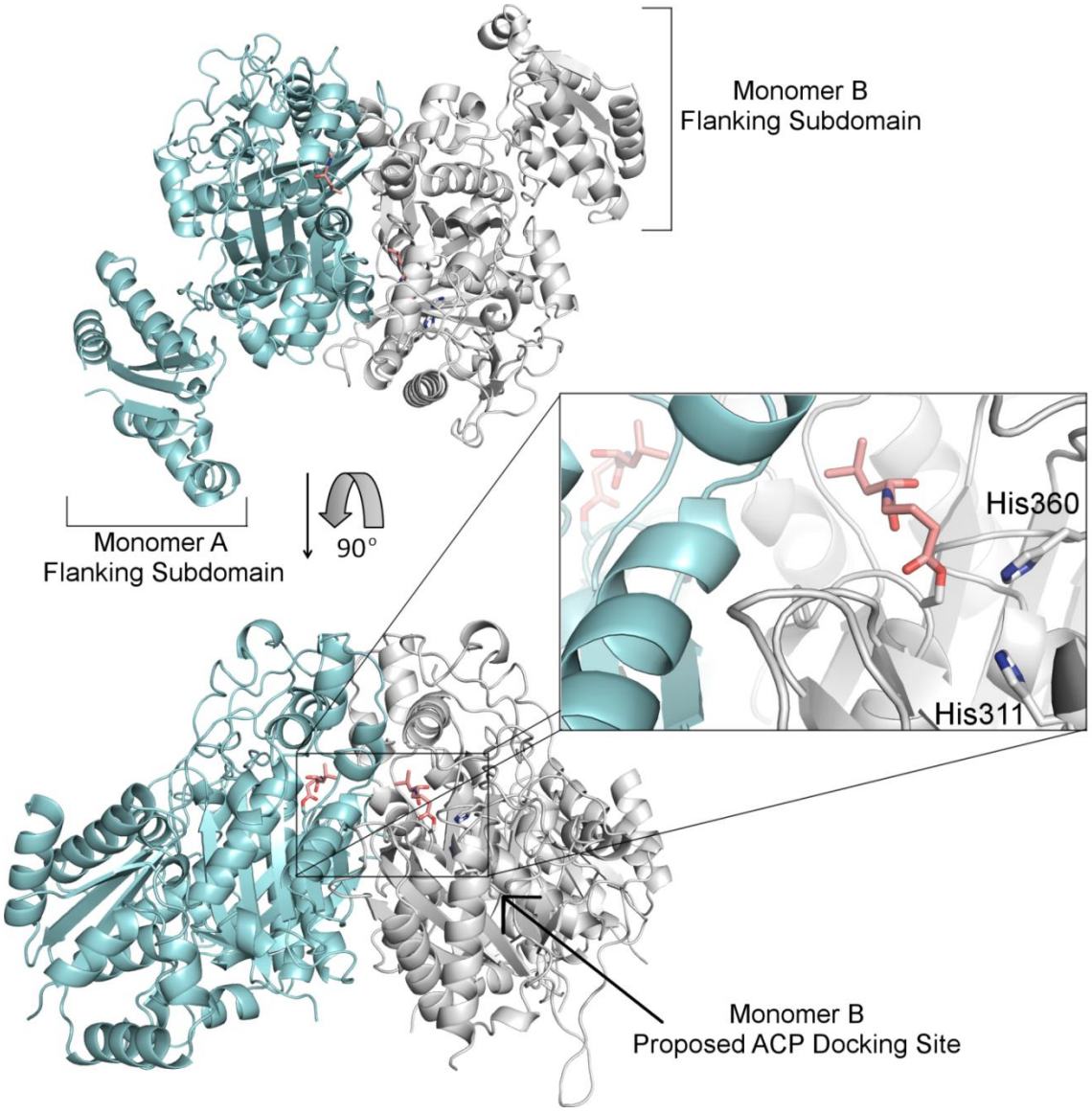


Figure 6-13. Location of substrate binding relative to proposed ACP docking site. Monomer A of PksKS2 is colored blue, monomer B is colored grey, and the covalently bound substrates are colored salmon (electron density was clearly defined for the substrate in both monomers of the complex structure). The proposed docking location for ACP has been boxed and expanded, with the catalytic histidines shown as sticks.

	PksKS2 Wild-type	PksKS2 C176S Natural Intermediate Complex	PksKS2 C176S Hexanoyl Complex
Data collection			
Space group	$P2_1$	C2	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	77.9, 113.1, 96.5	165.1, 71.3, 116.5	77.9, 113.3, 96.9
α, β, γ (°)	90.0, 107.0, 90.0	90.0, 95.1, 90.0	90.0, 106.9, 90.0
Resolution (Å)	30.7 – 1.90	43.5 – 2.30	33.7 – 2.83
R_{sym} or R_{merge}	0.085 (0.979)	0.151 (0.815)	0.184 (0.969)
$I / \sigma I$	23.7 (1.8)	6.3 (2.0)	11.8 (1.7)
Completeness (%)	97.4 (78.4)	93.3 (87.4)	98.5 (82.0)
Redundancy	3.6 (2.4)	3.6 (3.5)	7.4 (6.8)
Wilson B value (Å^2)	42.9	39.4	59.8
Refinement			
Resolution (Å)	29.70 – 1.95	43.47 – 2.35	33.08 – 2.90
No. reflections	114754	52829	35693
R_{work} / R_{free}	0.194 / 0.227	0.203 / 0.255	0.193 / 0.235
No. atoms			
Protein	9270	9161	9270
Ligand	-	28	14
Sulfate	30	-	30
Water	634	243	110
B-factors			
Protein	42.5	39.6	59.8
Ligand	-	28.5	53.2
Sulfate	63.9	-	92.9
Water	46.6	35.8	45.7
R.m.s. deviations			
Bond lengths (Å)	0.0124	0.003	0.0026
Bond angles (°)	1.298	0.699	0.701
Ramachandran Plot			
Favored (%)	97.4	96.9	96.4
Allowed (%)	2.6	3.1	3.6
Disallowed (%)	0.0	0.0	0.0

Values in parentheses are for highest-resolution shell.

Table 6-1. Crystallographic data and refinement statistics.

	Residue						
	109	**169**	**249**	**259**	**261**	**349**	**402**
trans-AT	Y	S	K	D	I	C	F
cis-AT	I	A	E	H	V	L	V

Table 6-2. Residues that consistently differ between cis- and trans-AT PKSs. These residues were found to be highly conserved within either *cis*- or *trans*-AT PKSs, but not between the two systems.

Chapter 7. The LINKS motif zippers *trans*-acyltransferase polyketide synthase assembly lines into a biosynthetic megacomplex

ABSTRACT

Polyketides such as the clinically-valuable antibacterial agent mupirocin are constructed by architecturally-sophisticated assembly lines known as *trans*-acyltransferase polyketide synthases. Organelle-sized megacomplexes composed of several copies of *trans*-acyltransferase polyketide synthase assembly lines have been observed by others through transmission electron microscopy to be located at the *Bacillus subtilis* plasma membrane, where the synthesis and export of the antibacterial polyketide bacillaene takes place. In this work we analyze ten crystal structures of *trans*-acyltransferase polyketide synthases ketosynthase domains, seven of which are reported here for the first time, to characterize a motif capable of zippering assembly lines into a megacomplex. While each of the three-helix LINKS (**L**ocalization **IN**ducing **K**etosynthase **S**equence) motifs is observed to similarly dock with a spatially-reversed copy of itself through hydrophobic and ionic interactions, the amino acid sequences of this motif are not conserved. Such a code is appropriate for mediating homotypic contacts between assembly lines to ensure the ordered self-assembly of a noncovalent, yet tightly-knit, enzymatic network. LINKS-mediated lateral interactions would also have the effect of bolstering the vertical association of the polypeptides that comprise a polyketide synthase assembly line.

INTRODUCTION

Complex polyketides are a class of secondary metabolites that provide a competitive advantage for a wide spectrum of bacteria and fungi, often by targeting vital processes within competing organisms (Hertweck, 2009; Piel, 2010). The polyketide synthases (PKSs) responsible for their production are inherently modular, enabling PKS-harboring organisms to explore vast regions of chemical space through the reorganization of enzymatic domains. While the clinical value of the antibiotics erythromycin and mupirocin allude to the potential of polyketides in drug discovery, harnessing the biosynthetic capabilities of PKSs has been hindered by the complexity of their higher-order architecture (Walsh, 2008; Keatinge-Clay, 2012; Dutta *et al*., 2014).

An enzymatic juggernaut of the subcellular milieu, a PKS typically contains upwards of 50 independently-folded domains, each responsible for one step in the biosynthesis of a polyketide. The enzymatic modules of PKSs work together in assembly-line fashion to extend and process a growing polyketide chain (Khosla *et al.*, 2014), each module containing a set of domains responsible for the incorporation and tailoring of a selected α-carboxylated extender unit. The domains employed by Type I PKSs are homologous to those responsible for fatty acid biosynthesis (Smith & Tsai, 2007), including acyltransferases (ATs) that select extender units, ketosynthases (KSs) that condense extender units with growing polyketides in a decarboxylative thio-Claisen condensation, ketoreductases (KRs) that stereospecifically reduce newly incorporated β-keto groups, dehydratases (DHs) that eliminate water to form α,β-olefins, enoylreductases (ERs) that stereospecifically reduce those olefins, and acyl carrier proteins (ACPs) that

shuttle polyketide intermediates between the domains within a module as well as the following module. Thus, modules containing different sets of KR, DH, or ER domains yield different functional groups and stereochemistries. A thioesterase (TE) downstream of the final module commonly off-loads the polyketide product via hydrolysis or cyclization.

Modular PKSs are divided into two architecturally-distinct classes: *cis*-AT and *trans*-AT (Piel, 2010). While *cis*-AT PKSs harbor ATs that are integrated into the multi-domain polypeptide, *trans*-AT systems rely on discretely-encoded AT domains that noncovalently interact with the megasynthase (**Figure 7-1**). Each of the independently-folded domains from *cis*-AT PKSs has been structurally characterized (Keatinge-Clay, 2012), and although structural information has recently become available for domains from *trans*-AT PKSs, if and how the eponymous *trans*-AT domains dock to the megasynthase to charge ACPs with extender units remains to be determined (Bretschneider *et al*., 2013; Gay *et al*., 2014; Piasecki *et al.*, 2014). A conserved ~100-residue region C-terminal to KS was hypothesized to facilitate the docking of *trans*-ATs to megasynthases and was named the AT-docking domain, or ATd (Tang *et al.*, 2004). The KS/AT adapters, observed in the crystal structures of KS+AT didomains from the erythromycin PKS and the mammalian fatty acid synthase (FAS), are homologous to ATd on a sequence level and seem to validate this functional assignment (Tang *et al.*, 2007; Maier *et al.*, 2008). However, *in vitro* experiments examining extender unit transfer from *trans*-ATs to ACPs showed that the inclusion of ATd in constructs decreased the

catalytic efficiency of acyl transfer compared to constructs that did not include ATd (Aron *et al.*, 2007).

The first-discovered, and prototypical, *trans*-AT PKS is encoded within the *Bacillus subtilis* and *Bacillus amyloliquefaciens* genomes (*pksX* and *bae*, respectively) (Piel, 2010). This hybrid PKS/non-ribosomal peptide synthetase (NRPS) produces bacillaene, a linear polyene diamide that interferes with prokaryotic protein synthesis (**Figure 7-1**) (Patel *et al.*, 1995). Transmission electron micrographs of *B. subtilis* cells have revealed dense masses associated with the plasma membrane, each comprised of several copies of the PksX megasynthase (Butcher *et al.*, 2007; Straight *et al.*, 2007). This organelle-sized structure, with an approximate mass of 10-100 MDa, includes the five multidomain polypeptides PksJ, PksL, PksM, PksN, and PksR. At the time of this discovery megacomplex formation was a complete mystery since *trans*-AT PKS polypeptides lack the N- and C-terminal docking domains that linearly organize *cis*-AT PKSs and no PKS architectural scaffolding proteins were known; however, ATd was hypothesized to help stabilize the observed structure (**Figure 7-1**) (Broadhurst *et al.*, 2008). Membrane localization of PKS enzymes has been observed for the mycolactone and actinorhodin pathways, and the biosynthetic machinery for the siderophore pyoverdine is known to accrete into a dense, membrane-bound mass termed the "siderosome"; the co-localization of large enzymatic complexes at the bacterial plasma membrane is emerging as a theme in the biosynthesis of natural products (Xu *et al.*, 2008; Imperi & Visca, 2013; Porter *et al.*, 2013; Simunovic *et al.*, 2006; Hantash & Earhart, 2000).

The first ATd to be structurally observed, from a didomain comprised of a KS and a β-branching enzyme (KS+B) from the rhizoxin PKS, revealed that ATd is structurally related to the KS/AT adapter of *cis*-AT PKSs (Bretschneider *et al*., 2013). Where the AT domain branches from the KS/AT adapter in *cis*-AT PKSs, only a short loop is present. The second structure of an ATd, from the second KS domain of the *B. subtilis* PksX synthase (PksKS2), showed two additional α-helices where the AT domain branches from the KS/AT adapter in *cis*-AT PKSs (Gay *et al*., 2014). Modeling a *trans*-AT docked to the ATd based on the location of AT within *cis*-AT PKSs indicated that significant restructuring of these helices would be required for docking. Since neither functional nor structural studies supported the role of ATd as a *trans*-AT docking site, it was renamed the "flanking subdomain".

Here, we present seven crystal structures of *trans*-AT KS domains that provide evidence that the flanking subdomain plays a role in the assembly of PKS megacomplexes. At first considered to be a serendipitous crystal contact, the clearly defined self-association of flanking subdomains published here and elsewhere appears in ten crystal structures (five different KS domains from three different microbial species harboring *trans*-AT PKSs; seven structures reported here, two structures reported in reference 9, and one deposited but not published structure, PDB: 4TL2). We introduce the term **L**ocalization **IN**ducing **K**etosynthase **S**equences (LINKS) to distinguish the three helices that form the observed interaction from the remainder of the flanking subdomain. Bioinformatic analysis of LINKS regions reveals that each is unique to the KS domain possessing them and indicates that PKS assembly lines that form

megacomplexes are zippered together by LINKS-containing KSs that possess higher affinity for copies of themselves than other LINKS-containing KSs within the assembly line. These findings suggest an unprecedented mechanism for the assembly of several copies of a PKS megasynthase into a megacomplex.

RESULTS

Structure of the LINKS interaction

Three of the *trans*-AT KS domains deposited into the Protein Data Bank (PDB) contain the LINKS region: one from the bacillaene PKS (PksKS2, PDBs 4NA1 and 4NA2) and two from the migrastatin PKS (MgsKS5 and MgsKS7, PDBs 4TL2 and 4TKT, respectively). The LINKS interaction between homodimeric KSs, not apparent from the asymmetric units of wild-type PksKS2 (PDB 4NA1) and MgsKS5 (PDB 4TL2), can be observed by generating the crystallographic symmetry mates. This is opposite in crystals of PksKS2(Cys176Ser) (PDB: 4NA2), in which the asymmetric unit shows two KS monomers associated via the LINKS interaction and the KS dimer is visualized by generating crystallographic symmetry mates. Here we report seven additional crystal structures in which LINKS interactions are observed, including one of PksKS2 in its monomeric form, one of PksKS6, two of PksKS6(Cys169Ser), one of Pks(ACP5+KS6), one of BaeKS1, and one of BaeKS5 (**Table 7-1**). To date, only one crystal structure of a LINKS-containing *trans*-AT KS has been reported that does not show this interaction (MgsKS7, PDB: 4TKT).

The LINKS motif is comprised of ~40 residues (α17-α19) that project from the flanking subdomain as a triangular flap (see reference 9 for secondary structure assignments) (**Figure 7-2**). While α19 is structurally equivalent to a helix in the KS/AT adapter of *cis*-AT PKSs and the mammalian FAS, α17 and α18 take the place of the AT domain in these synthases. Helices α17 and α18 do not make interactions with neighboring secondary structural elements and thus are not rigidly oriented with respect to the flanking subdomain. In the LINKS interaction α18 of one flanking subdomain makes contact with a spatially-reversed α18 of the neighboring flanking subdomain, and α17 of one flanking subdomain makes contact with α19 of the neighboring flanking subdomain.

The observed LINKS interactions form through the burial of hydrophobic residues and exhibit a high degree of shape complementarity (~700 $Å^2$). In the LINKS motif of PksKS2 (the structure possessing the highest-resolution and lowest B-factors for the LINKS region) these hydrophobic residues include I552, F553, M559, L563, W566, and L578 (**Figure**s **7-2b-d** and **Table 7-2**). The LINKS interactions observed from bacillaene synthases (PksKS2, PksKS6, BaeKS1, and BaeKS5) show contact between clusters of charged residues that cap the poles of α18. The N-terminal end of this helix contains ~3 negatively-charged residues, and the C-terminal end contains ~3 positively-charged residues. Two copies of α18 are oriented in an antiparallel fashion, and the negatively-charged N-terminal end of α18 from one flanking subdomain forms salt bridges with the positively-charged C-terminal end of α18 from the neighboring flanking subdomain (**Figure 7-3**). Collectively, the burial of exposed hydrophobic residues in the

center of each LINKS triangular flap and the clustering of complementary ionic interactions at both termini of α18 form a lock-and-key fit uniquely adapted to each LINKS interaction (**Figure 7-4**).

The LINKS motifs connect neighboring KS dimers in crystals of MgsKS5 (PDB: 4TL2); however, instead of the twofold axes of these KS dimers being parallel as in each of the other observed LINKS-mediated associations of KS dimers, they are nearly orthogonal (**Figure 7-2a**). The MgsKS5 LINKS interaction is distorted relative to all the other crystallographically-observed LINKS interactions - α17 and α18 are skewed, and several ionic interactions potentially made *in vivo* are not formed (e.g., Asp1121/Arg1133 and Glu1122/Arg1129) (**Figure 7-5**). The only intact salt bridge, formed by Lys1120 and Asp1138, may provide an example of covariation within the LINKS motif, as both of these residues generally carry the opposite charge in sequence alignments (**Figure 7-6** and **Table 7-3**).

The seven crystal structures reported here show the LINKS interaction in seven different contexts. In one crystal form, PksKS2 monomers that do not make the traditional KS homodimer interaction were observed. Remarkably, even though ~5300 $Å^2$ is not buried through the formation of the traditional KS homodimer, the LINKS interaction is maintained (**Figure 7-2a**). The construct Pks(ACP5+KS6), generated to investigate the interaction between KS and ACP, also yielded crystals in which the LINKS interaction was observed. Although the crystals could not be optimized to diffract beyond 4.0-Å resolution and the ACP domain could not be located in the electron density map, the LINKS interaction was observed to be formed by each of the eight KS

monomers in the asymmetric unit (**Figure 7-7**). In addition to the 2.0 Å-resolution structure of PksKS6(Cys169Ser), a construct generated to visualize bound intermediates (Gay *et al*., 2014), two additional crystal forms of PksKS6 were observed that contain the LINKS interaction (wild-type PksKS6, 3.10-Å resolution; PksKS6(Cys169Ser), 2.16-Å resolution) (**Table 7-1**). Two crystal structures of KSs from the *B. amyloliquefaciens* bacillaene PKS, BaeKS1 (2.93 Å) and BaeKS5 (4.20 Å), were also determined. While locally-elevated temperature factors prevented the complete modeling of α17-α19 in BaeKS1 (**Figure 7-8**), the relative positioning of KSs indicates that each of the flanking subdomains form LINKS interactions.

Bioinformatic analysis of LINKS

The lack of sequence conservation of α17-α19 is a hallmark of the LINKS motif. While the LINKS motifs shown in **Figure 7-6** are each quite hydrophobic and contain charged residues at the poles of α18, individual LINKS residues are highly variable (**Figure**s **7-6** and **7-9**). To further investigate this variability the sequence conservation of LINKS-containing KSs was mapped onto a model of a *trans*-AT KS domain (**Figure 7-10**). An alignment of 152 *trans*-AT KS sequences was generated, and the degree of conservation of each residue was assessed. A moving average calculation was used to determine the mean conservation for short stretches of sequence, such that the value for the conservation of residue *n* is equal to the average value of conservation for the 11 residues ranging from *n*-5 to *n*+5. A template KS structure was then colored based on this value: red for 0-25%, yellow for 25-50%, green for 50-75%, and blue for 75-100%.

While the active site region of KS appears blue and the remainder of the KS body varies between green and yellow, the longest stretch of red (46 residues) is the LINKS region. The next longest stretch of red sequence is only 17 residues and is located on the face of KS that would be adjacent to processing domains such as DH, KR, and MT. Classes of LINKS motifs other than the one shown in the sequence alignment of **Figure 7-6** likely exist; several LINKS motifs that did not align well with those in **Figure 7-6** are displayed in **Figure 7-9**.

LINKS interactions are homotypic

With the exception of MgsKS5, the relative orientation of *trans*-AT KS dimers connected through a LINKS interaction is essentially identical for all of the crystal structures available to date (**Figure 7-2a**). We investigated whether heterotypic LINKS interactions were complementary between KSs from different modules. The six LINKS helices (three helices from each KS monomer) were isolated from the observed LINKS interactions (using the structure with the best-resolved LINKS motif for each KS), each of these were superposed, and new coordinate files were created for the heterodimeric LINKS interactions (i.e., PksKS2 with BaeKS1, PksKS6, or MgsKS5). The simulated interactions were investigated to assess the interfacial contacts. In general, heterotypic LINKS interactions were significantly less favorable than homotypic LINKS interactions, displaying much worse shape complementarity with fewer ionic interactions (**Figure 7-11**).

DISCUSSION

The docking site of *trans*-ATs had been hypothesized to be generated by an ~100-residue subdomain, formerly termed ATd and now referred to as the "flanking subdomain", that is associated with the majority of KSs within *trans*-AT PKSs. That an AT domain is integrated into the equivalent subdomain of the mammalian FAS and *cis*-AT PKSs appears to support this role; however, this subdomain is common to each of these synthases and was most likely present within an ancestral PKS before *cis*- and *trans*-AT PKSs diverged (Nguyen, *et al*., 2008). Thus, the docking site to which *trans*-ATs dock in *trans*-AT PKSs should not be expected to be equivalent to where ATs are integrated in *cis*-AT PKSs. A ~40-residue LINKS motif is usually found in *trans*-AT PKSs where the AT domain is integrated in *cis*-AT PKSs. In 10 out of 11 crystal structures of constructs containing the LINKS motif it is found mediating a LINKS-interaction, and in 8 of the 10 structures it associates neighboring KS dimers in a parallel orientation (the exceptions being MgsKS5 and the monomeric form of PksKS2). We propose that the LINKS motif makes lateral interactions that help construct megacomplexes like those observed in *B. subtilis* cells (**Figure 7-12**; based on the measured dimensions, one bacillaene megacomplex is comprised of 10-20 homodimeric bacillaene PKS assembly lines).

The LINKS hypothesis excludes the site that AT is observed in *cis*-AT PKSs as the docking site for a *trans*-AT. While *trans*-ATs could dock with another region of the flanking subdomain or the KS body, they may only interact with the ACP domains to which they transfer extender units (Ye *et al.*, 2014). Studies have shown that PksE, a

trans-AT in the PksX bacillaene pathway, is present in concentrations 10–100 fold greater than the assembly line polypeptides (Straight *et al.*, 2007). Additionally, superstoichiometric *trans*-AT concentrations have a large effect on the rate of polyketide production by the virginiamycin *trans*-AT PKS (Pulsawat *et al.*, 2007). If *trans*-ATs were stably associated with each *trans*-AT PKS module, increased *trans*-AT expression would have little effect. Several condensation-incompetent *trans*-AT KSs harbor flanking subdomains yet have no need to interact with a *trans*-AT (e.g., BTKS10, ChiKS18, DszKS10, and OnnKS4). Similarly, several *trans*-AT KSs that catalyze polyketide chain elongation do not contain a flanking subdomain, yet the corresponding ACPs from these modules must interact with a *trans*-AT (e.g., ChiKS16, LnmKS3, MlnKS8, and DifKS10). Studies have shown that *trans*-ATs are capable of charging ACP domains in the absence of a KS or flanking subdomain and that a *trans*-AT from the kirromycin PKS (KirCII) is quite specific for its cognate ACP (Aron *et al.*, 2007; Ye *et al.*, 2014; Wong *et al.*, 2011; Lopanik *et al.*, 2008; Musiol *et al.*, 2011). This degree of *trans*-AT specificity for an ACP is surprising if the flanking subdomain or KS body is selective for docking with a particular *trans*-AT.

The structure of the homodimeric KS+B didomain from the rhizoxin *trans*-AT PKS shows that the longest dimension of the B dimer is rotated 90˚ relative to the longest dimension of the KS dimer (Bretschneider *et al.*, 2013). This orientation, also anticipated for the *trans*-AT KS+DH didomain based on the structural homology of the B and DH domains, would be compatible with the higher-order architecture of a PKS megacomplex formed through LINKS interactions (**Figure 7-12**). While several contacts between the

KS and B dimers from a *trans*-AT PKS fix their relative orientation, the KS and DH dimers of the mammalian FAS rotate freely relative to one another (Maier *et al.*, 2008; Brignole *et al.*, 2009). The crystal structure of the mammalian FAS shows no contact between the two dimers (**Figure 7-13**), and cryo-electron microscopy studies also report flexibility between the KS and DH dimers.

Even if LINKS motifs do not make high affinity interactions, significant avidity could be generated through the several LINKS motifs interacting along the length of the synthase (**Figure 7-14**). Within the megacomplex, the lateral LINKS interactions would have the effect of bolstering vertical interactions, thought to be mediated by the N- and C-terminal regions of *trans*-AT PKS polypeptides. The flexibility of the LINKS interaction, illustrated by the MgsKS5 crystal structure, could aid megacomplex formation and maintenance. Many *trans*-AT PKS assembly lines harbor several LINKS motifs (e.g., the bacillaene PKS contains 12, the chivosazol PKS contains 9, the mupirocin PKS contains 7), such that the dissociation of a single LINKS interaction would have little effect on the stability of the entire megacomplex. LINKS interactions may be difficult to biophysically characterize using methods other than x-ray crystallography. If the dissociation constant for an individual LINKS interaction is in the millimolar range, it may not be observable through techniques such as small-angle x-ray scattering (SAXS) or analytical ultracentrifugation (Davison *et al.*, 2014). To accurately determine the dissociation constant of the LINKS interaction, the natural KS dimerization interface may need to be disrupted by site-directed mutagenesis such that only the LINKS-mediated association of KS monomers is measured.

Several structures of *trans*-AT PKS KS domains have been deposited in the PDB that do not contain a LINKS motif, including the KS+B didomain from the rhizoxin PKS (Bretschneider *et al*., 2013). However, only one reported crystal structure of a LINKS-containing KS does not display the LINKS interaction (MgsKS7). Favorable crystal contacts made elsewhere may preclude this interaction in the same manner that LINKS-mediated crystal contacts within the monomeric crystal structure of PksKS2 are favored over natural homodimer interactions.

The mode through which LINKS interactions stabilize *trans*-AT PKS megacomplexes has not been observed in the architectures of other biosynthetic assemblies. The crystal structures presented here reveal a defined binding interface formed by hydrophobic and ionic interactions mediated by three helices on the surface of the flanking subdomain. Bioinformatic analysis suggests that the LINKS interaction is unique for each KS, such that the in-register, homotypic, lateral interactions made by them also confer stability to the vertical interactions between the assembly lines of the megacomplex through avidity. While *trans*-AT systems represent an excellent template for the rational engineering of synthetic PKSs for the exploration of new medicines, it may be necessary to consider the implications of modifying the LINKS network when genetically relocating modules or domains within a *trans*-AT PKS megasynthase.

MATERIALS AND METHODS

All protein structures described in this work were cloned, purified, and the corresponding structures refined using identical methods to those previously described

for PksKS2 (Gay *et al*., 2014). Therefore, only the primers used for gene amplification of new constructs, crystallization conditions, phasing methods, and any other modifications will be reported here.

Crystals of PksKS2 in which the KS domains did not crystallize in the native dimeric form grew over a period of 5-7 days by sitting drop vapor diffusion at 4 ˚C. Drops were formed by mixing 2 µL protein solution (4.5 mg/ml PksKS2, 150 mM NaCl, 10% glycerol, 10 mM HEPES, pH 7.5) with 1 µL crystallization buffer (30% PEG 400 (v/v), 0.2 M $MgCl_2$, 0.1 M HEPES, pH 7.5). Crystals were frozen in liquid nitrogen after a 20 min soak in crystallization buffer modified with 20% glycerol, and the diffraction data collected at ALS Beamline 5.0.3. The structure was solved to 3.0 Å resolution by molecular replacement with PhaserMR in the CCP4 suite, using a single chain of the dimeric structure of PksKS2 (PDB: 4NA1) as the search model.

The DNA encoding PksKS6 was amplified using primers: 5′-GCGGCCTGGTGCCGCGCGGCTCTAGCGACCGCCCGGAGGATGCGATAG-3′ and 5′-GTGGTGGTGGTGGTGGTGATGTTAGCCTTCTTGAGTTGGCAGCCAG-3′ (LIC cloning regions underlined for insertion into the plasmid pGAY28 (Gay *et al.,* 2014)). The eluted protein was concentrated to 15 mg/mL in the equilibration buffer and stored at -80℃ until needed. Crystals of PksKS6 grew in 1-6 days by sitting drop vapor diffusion at 22 ˚C. Drops were formed by mixing 2 µL protein solution (15 mg/mL PksKS6, 150 mM NaCl, 10 mM HEPES, pH 7.5) with 1 µL crystallization buffer (150 mM $(NH_4)_2SO_4$, 15% PEG 4000 (v/v), 100 mM MES, pH 6.0). Crystals were frozen in liquid nitrogen after a 20 min soak in the crystallization buffer modified with 10% (v/v) ethylene glycol.

Diffraction data, collected at ALS Beamline 5.0.3, were processed by HKL2000. The structure was solved to 3.00 Å resolution by molecular replacement using the KS2 monomer from the PksX synthase of *B. subtilis* as the search model (PDB: 4NA1).

The PksKS6(Cys169Ser) expression plasmid was generated by PCR amplification from the PksKS6 construct described above using primers 5′-AGACGGCCTCGAGCTCTTCATTGGTCGGCACTCATTTAGCGCGCCAGGCACTTATAAACAAAG-3′ and 5′-AGAACTGCTCGAGGCGGTATCAATAGGAATCGCTGGCCCTTTTAAATTTAGAAAATACGG-3′ (XhoI site underlined). The PCR product was digested with XhoI and ligated. Crystals of PksKS6(Cys169Ser) grew over 2-10 days by sitting drop vapor diffusion at 22 ˚C. Drops were formed by mixing 2 μL protein solution (15 mg/mL PksKS6(Cys169Ser), 150 mM NaCl, 10 mM HEPES, pH 7.5) with 1 μL crystallization buffer (1.7 M $LiSO_4$, 100 mM Tris, pH 8.0) (both crystal forms grew in the same crystallization condition). Crystals were frozen in liquid nitrogen after a 20 min soak in crystallization buffer modified with 10% (v/v) ethylene glycol and the diffraction data were collected at APS Beamline 23-ID-D. The structure was solved by molecular replacement using a single chain of the dimeric structure of PksKS2 (PDB: 4NA1) as the search model.

Bacillus amyloliquefaciens FZB42 genomic DNA was used as template DNA to PCR amplify BaeKS1 with KAPA High-Fidelity DNA polymerase (KAPA) using the forward primer 5′-

GCGGCCTGGTGCCGCGCGGCTCTAGCTTTCCCGATTATTACGAGGACAG-3′ and the reverse primer 5′-GTGGTGGTGGTGGTGGTGATGTTAGTTCTTTTCTGCTTTTTTCGGCG-3′ (pGAY28 LIC cloning regions underlined). The eluted protein was concentrated to 14 mg/mL in the equilibration buffer and stored at -80℃ until needed. Crystals of BaeKS1 grew over 4-15 days by sitting drop vapor diffusion at 22 ˚C. Drops were formed by mixing 2 µL protein solution (14 mg/mL BaeKS1, 150 mM NaCl, 10 mM HEPES, pH 7.5) with 1 µL crystallization buffer (32% PEG 400 (v/v), 0.2 M $MgCl_2$, 100 mM HEPES, pH 7.5). Crystals were frozen in liquid nitrogen after a 20 min soak in crystallization buffer modified with 40% PEG 400 (v/v) and the diffraction data were collected at APS Beamline 23-ID-D. The structure was solved by molecular replacement using a single chain of the dimeric structure of PksKS2 (PDB: 4NA1) as the search model.

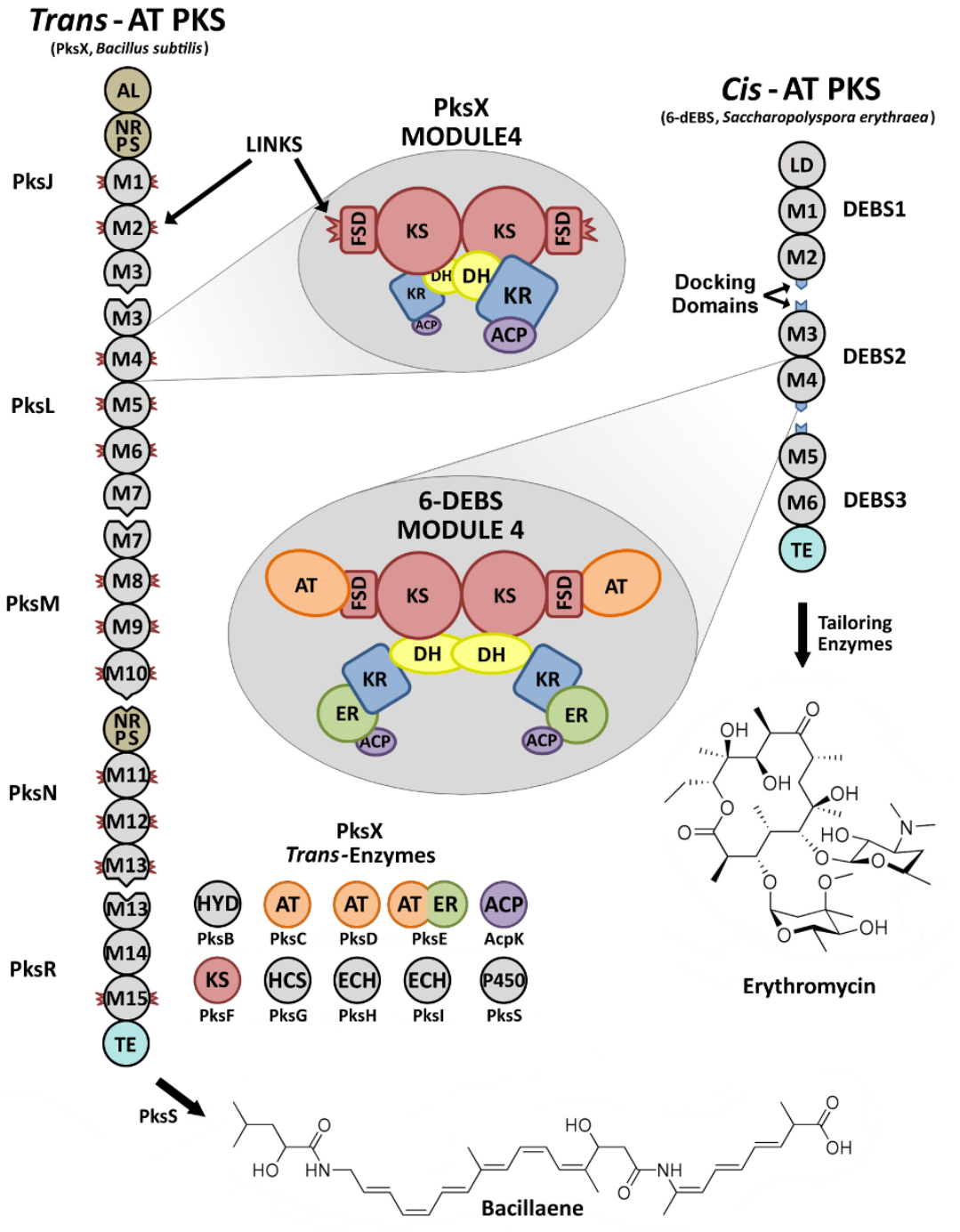
The gene corresponding to Pks(ACP5+KS6) was PCR amplified using primers 5′-GCGGCCTGGTGCCGCGCGGCTCTAGCGCTGAAGAAACGATTCAATATGC-3′ and 5′-GTGGTGGTGGTGGTGGTGATGTTAGCCTTCTTGAGTTGGCAGC-3′ (pGAY28 LIC cloning regions underlined). Crystals grew over a period of 2-4 days by sitting drop vapor diffusion at 22 ˚C. Drops were formed by mixing 2 µL protein solution (25 mg/ml Pks(ACP5+KS6), 150 mM NaCl, 10% glycerol, 10 mM HEPES, pH 7.5) with 1 µL crystallization buffer (30% PEG 4000 (v/v), 0.2 M sodium acetate, 0.1 M Tris, pH 8.5). Crystals were frozen in liquid nitrogen after a 20 min soak in crystallization buffer

modified with 20% glycerol, and the diffraction data collected at ALS Beamline 5.0.2. Phases were obtained by molecular replacement using PksKS6 (wild-type) as a search model. The structure was solved to 4.0 Å resolution by molecular replacement using a single chain from the dimeric structure of PksKS6 as the search model. Electron density for the eight ACPs in the asymmetric unit could not be located despite exhaustive inspection of the map.

The DNA encoding BaeKS5 was amplified using primers: 5′-GCGGCCTGGTGCCGCGCGGCTCTAGCCAGCAGCTGACAGAGCGTG-3′ and 5′-GTGGTGGTGGTGGTGGTGATGTTAATCAGCTGATGTGTCGATCCAATAAC -3′ (pGAY28 LIC cloning regions underlined). The eluted protein was concentrated to 9.9 mg/mL in the equilibration buffer. Crystals of BaeKS5 grew in 2 days by sitting drop vapor diffusion at 22 ˚C. Drops were formed by mixing 1.75 μL protein solution (9.9 mg/mL BaeKS5, 150 mM NaCl, 15 mM HEPES, pH 7.5) with 0.5 μL crystallization buffer (1.1 M $LiSO_4$, 0.1 M Tris, pH 8.0). Crystals were frozen in liquid nitrogen after a 20 min soak in the crystallization buffer modified with 10% (v/v) glycerol. Diffraction data, collected at APS Beamline 23-ID-B, were processed by HKL2000. The structure was solved to 4.20 Å resolution by molecular replacement using the KS2 monomer from the PksX synthase of *B. subtilis* as the search model (PDB: 4NA1).

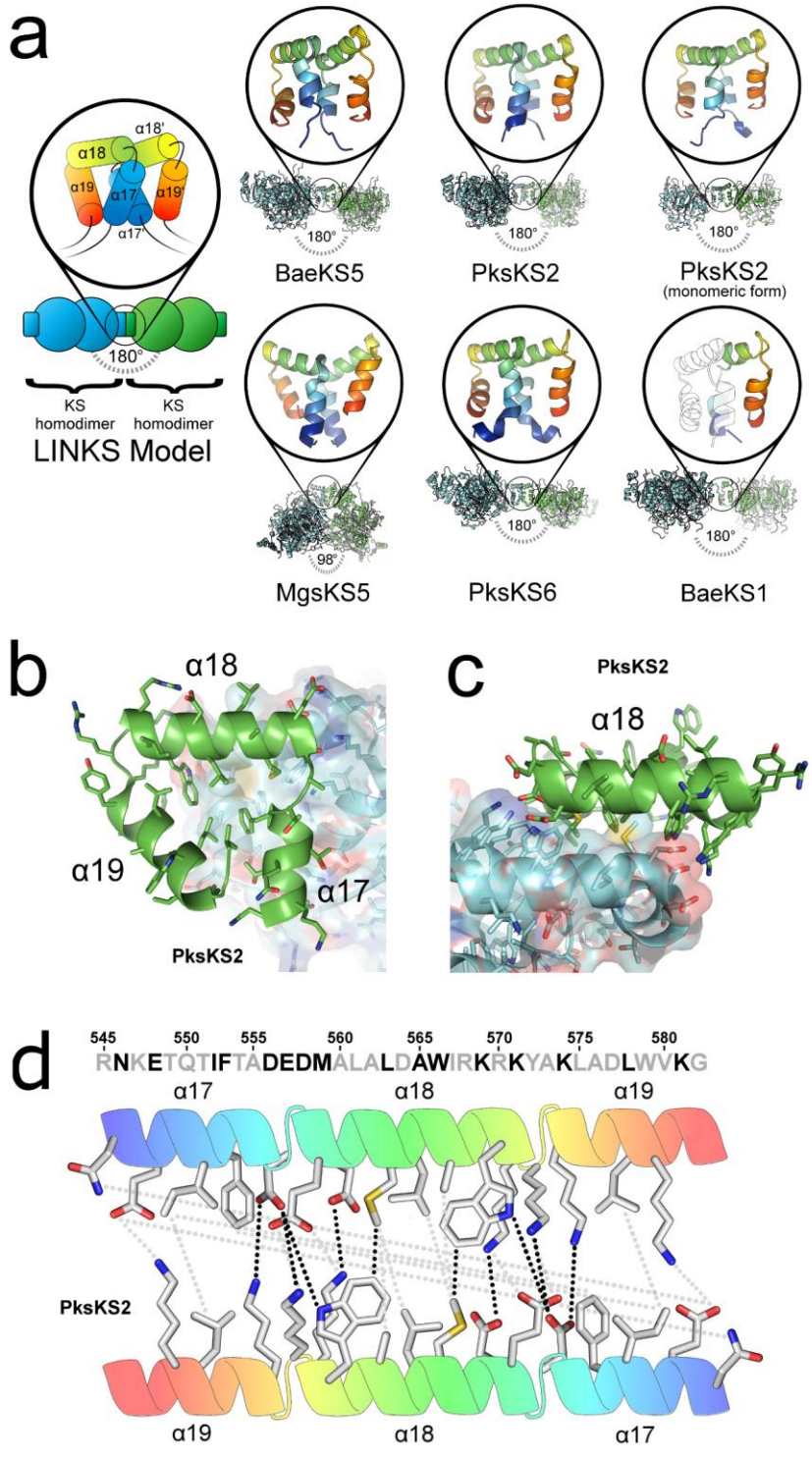
ACKNOWLEDGEMENTS

Instrumentation and technical assistance for crystallographic work were provided by Dr. A. Monzingo and the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health (NIH), National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. We thank the NIH (GM106112) and the Welch Foundation (F-1712) for supporting this research.



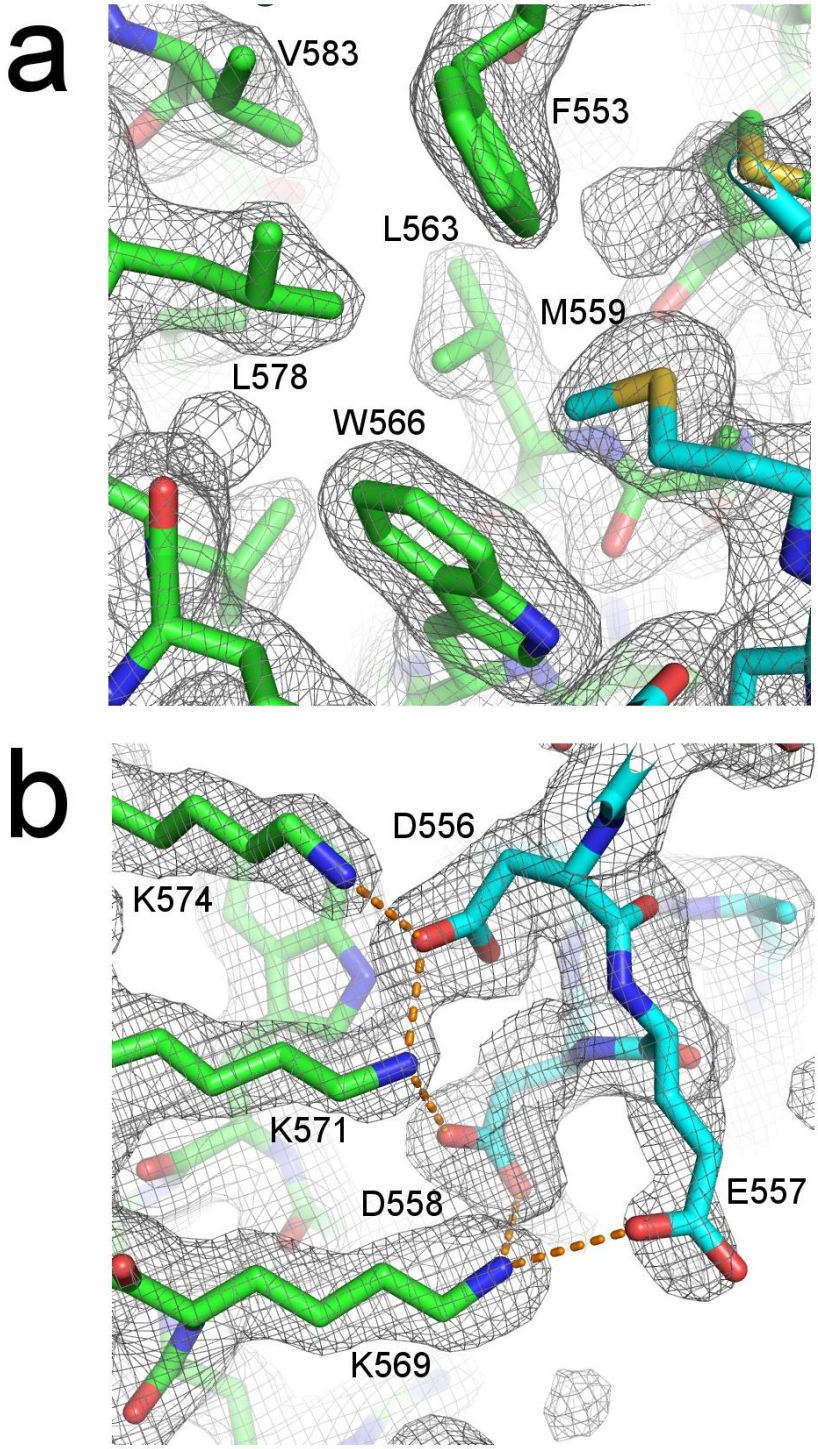
(Figure 7-1 continued on next page.)

Figure 7-1. *Trans*-AT vs. *cis*-AT PKS architecture. The modular assembly for the archetypal *trans*-AT (PksX) and *cis*-AT (6-dEBS) PKSs are shown, responsible for the biosynthesis of bacillaene and erythromycin, respectively. For each system, the fourth PKS module has been magnified to reveal the organization of domains. Although much of the higher-order architecture remains unknown, the schematics represent hypothetical models based on available crystal structures (Ery(KS+AT)3 (PDB: 2QO3), EryDH4 (PDB: 3EL6), Spn(KR+ER)2 (PDB: 3SLK), PksKS2 (PDB: 4NA1), and Rhi(KS+B)11 (PDB: 4KC5). The presence of a LINKS contact is shown as lateral red extensions for each of the modules in PksX that harbor this region. N- and C-terminal docking domains for 6-dEBS are shown as vertical blue extensions. ACP, acyl carrier protein; AL, acyl-CoA ligase; AT, acyltransferase; DH, dehydratase; ECH, enoyl-CoA hydratase; ER, enoylreductase; FSD, flanking subdomain; HCS, HMG-CoA synthase; HYD, hydrolase; KS, ketosynthase; KR, ketoreductase; LD, loading didomain; M, module; NRPS, non-ribosomal peptide synthetase; P450, cytochrome P450; TE, thioesterase.



(Figure 7-2 continued on next page.)

Figure 7-2. The LINKS interaction. The LINKS interaction that forms between *trans*-AT KS domains is mediated by three α-helices (α17-α19) that bind to a spatially-reversed copy of the same structure. **(a)** A model for the LINKS interaction between two KS homodimers is magnified, revealing the orientation of the LINKS helices. The angle above the structures indicates the relative angle between KS monomers forming the interaction (Figure 7-5). Each individual LINKS sequence is colored blue at the N-terminal end of α17 and red at the C-terminal end of α19, and the N-terminal end of α17 always forms favorable interactions with the C-terminal end of α19 from the neighboring KS (α19′). The coloring scheme used for the cartoon model has been repeated in each of the displayed structures to highlight the conservation of the LINKS interface. The structure shown for PksKS2 (monomeric form) reveals two KS monomers that do not form the traditional homodimer interaction; however, the LINKS interaction is crystallographically maintained. The relatively high thermal factors for the LINKS helices in BaeKS1 (3.0 Å) did not permit the complete construction of α17-α19 for each of the KS monomers, but the relative orientation of the KS bodies indicates that the LINKS interaction is conserved. **(b)** The three helices forming the LINKS interaction for PksKS2 are shown in cartoon format (green), and the remainder of this PksKS2 monomer has been hidden for clarity. The neighboring KS forming the complementary LINKS interaction is represented with a transparent surface (cyan). The angle of the image is set from the interior of the flanking subdomain, to reveal the collection of centrally-located hydrophobic residues that would be surface-exposed in the absence of the LINKS interaction. **(c)** A 90˚ rotation of the viewing angle shown in panel B reveals the series of ionic interactions formed at the poles of α18. The aspartate and glutamate residues at the N-terminal end of α18 form well-defined ionic interactions with the lysine residues at the C-terminal end of α18 from the neighboring structure. Met559 of the blue monomer can be observed in the center of the image, extending into the hydrophobic cavity of the green monomer (Figure 7-3a). **(d)** The three LINKS helices for PksKS2 have been modeled linearly to reduce the complexity of the image. LINKS contacts between residues ranging from 2.0 - 3.0 Å are shown as black dots, and those ranging from 3.1 – 4.0 Å are shown as grey dots. The corresponding sequence for the region is also shown, and residues represented by grey letters have been hidden from the cartoon and do not directly contribute to the LINKS interaction.



(Figure 7-3 continued on next page.)

Figure 7-3. Electron density for PksKS2 LINKS interaction. A $2F_o$-F_c map contoured at 1.0 r.m.s.d. is shown for the PksKS2 LINKS interaction, which represents the highest-resolution view available of this region. **(a)** M559 from the blue monomer inserts into the hydrophobic core of the opposite KS monomer (green). This interaction is mirrored several angstroms away, employing the same residues from the alternate KS. **(b)** The N-terminus of α18 (blue) forms well-defined ionic interactions with the C-terminus of α18 (green) from the complementary KS. This interaction is repeated on the opposite end of α18 for both helices, albeit with the correspondingly-exchanged residues. For specific interatomic distances, see Table 7-2.

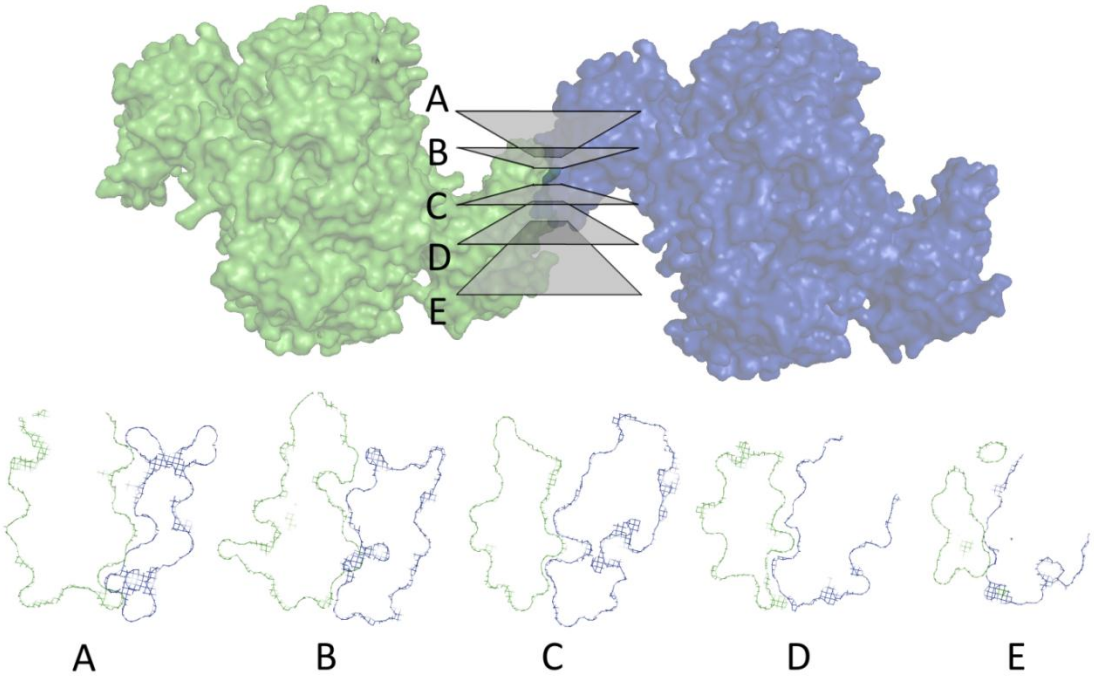


Figure 7-4. Shape complementarity of LINKS interaction. The hydrophobic core of the LINKS interaction exhibits a high degree of shape complementarity, maximizing the van der Waals forces that stabilize the contact. A crystallographically-observed LINKS interaction is shown (PksKS2), cut by five planes (A-E). Each resulting cross-section reveals that cavities or alcoves present in one KS are complemented by protrusions and extensions from the neighboring KS.

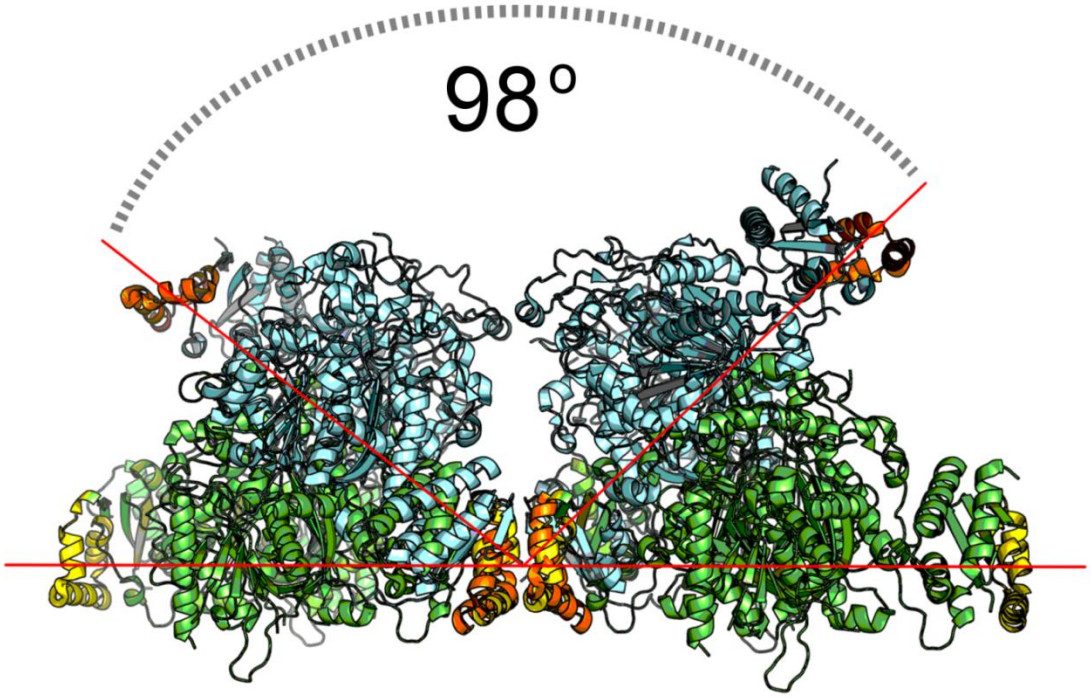
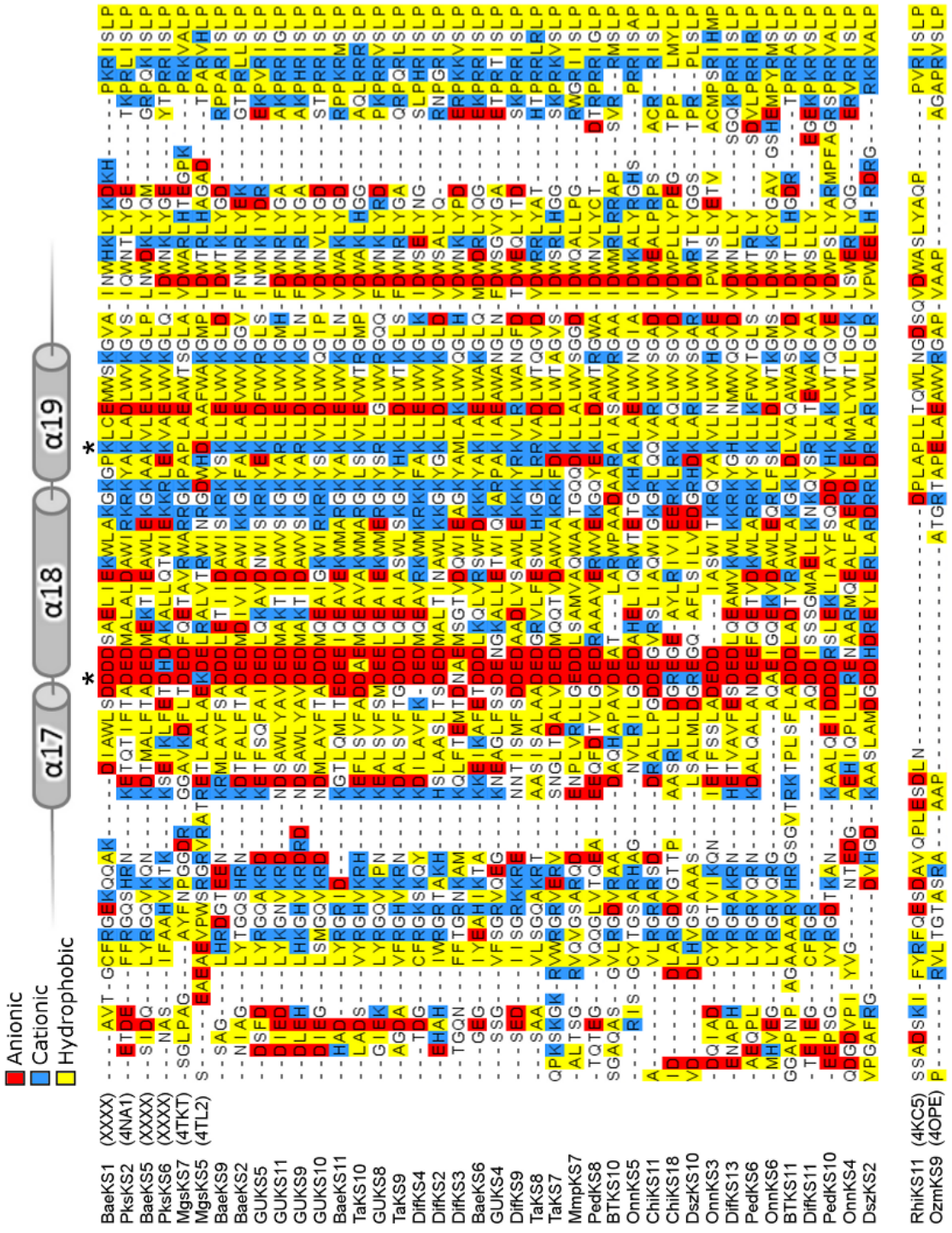


Figure 7-5. Angle of MgsKS5 LINKS interaction relative to PksKS2. Shown in green is PksKS2 with corresponding LINKS highlighted yellow, and shown in cyan is MgsKS5 with corresponding LINKS highlighted orange. Both structures were generated from the associated crystallographic symmetry operators. Structurally aligning both LINKS interactions reveals the flexibility of α17-α19, indicating that α18 is able to swing at least ~50˚ from the FSD to form the LINKS interaction. The angle for PksKS2 is 180˚, because a straight line can be drawn through the center of α17-α19 for each of the four monomers displayed. It is expected that 180˚ represents the lowest energy angle for the interaction, and deviations from this can arise from favorable contacts made elsewhere in the crystal.



(Figure 7-6 continued on next page.)

Figure 7-6. Sequence alignment of the LINKS region. Over 100 *trans*-AT KS sequences were aligned, and a subset of the alignment is shown. Residues in the alignment have been colored according to the following scheme: red for acidic residues, blue for basic residues, and yellow for hydrophobic residues. Helices α17-α19 (grey) delineate the boundaries of the LINKS region. If a KS in the alignment has been structurally characterized, the associated PDB identifier is shown in parentheses (XXXX: reported in this work, to be deposited). Although strict conservation of the LINKS region is poor, the conservation of LINKS attributes are generally maintained (i.e., charged poles of α18 flanked by hydrophobic residues). For those KSs that do not contain LINKS (e.g., RhiKS11 and OzmKS9, bottom of alignment), the available structures do not reveal any crystallographic homotypic interactions mediated by this region. The only structure available that contains LINKS but does not exhibit the crystallographic interaction observed in all other *trans*-AT KS structures is of MgsKS7. The structure of MgsKS5 reveals the LINKS interaction, yet the LINKS attributes differ from what has been observed in other structures. A seemingly anomalous basic residue at the N-terminal end of α18 (Lys1120) forms an ionic interaction with an acidic residue at the N-terminal end of α19 (Asp1138) that is almost invariably basic in the other sequences (both residue positions are marked with asterisks). (A 22-residue region from MgsKS5 that is composed of alternating aspartate and serine residues has been deleted from the image for clarity, identified by "†").

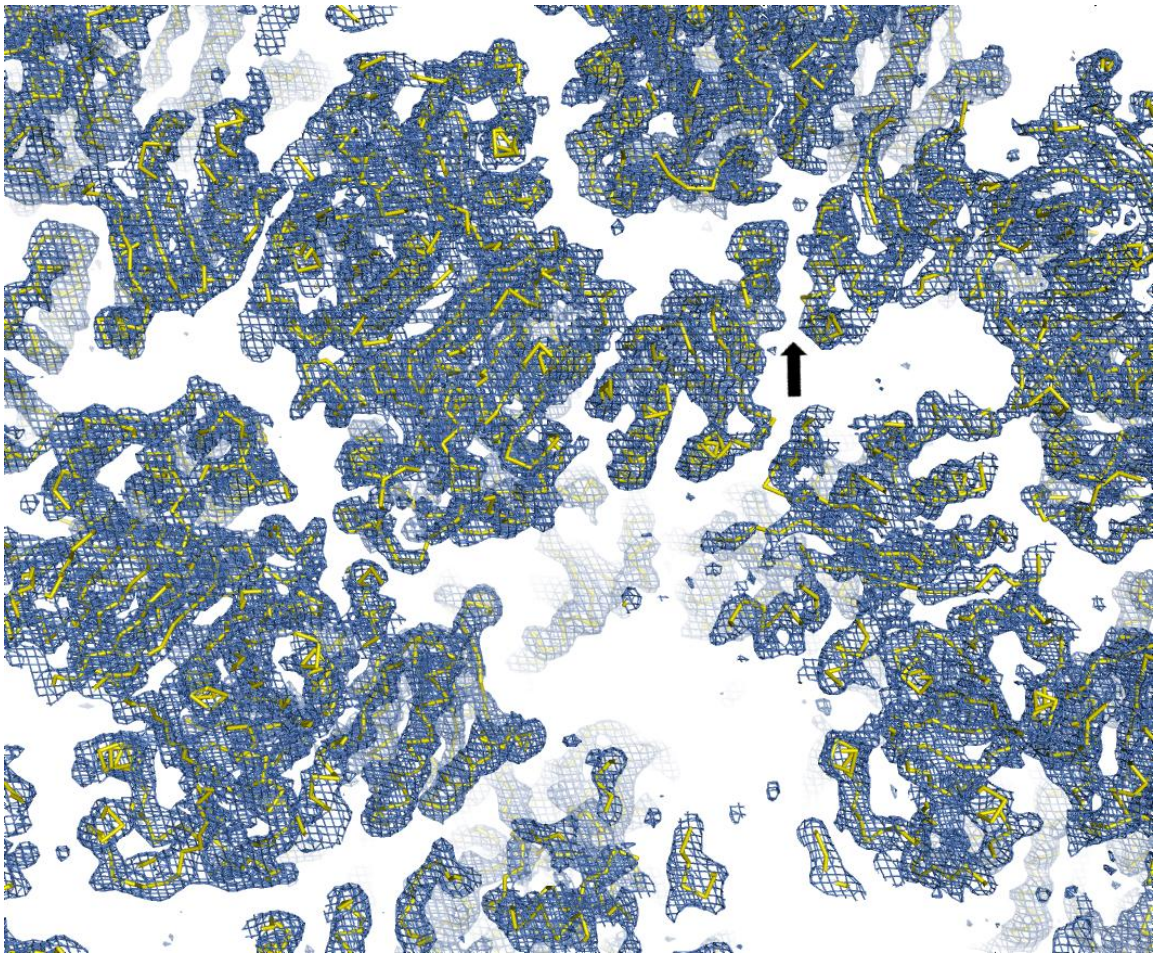


Figure 7-7. $2F_o$-F_c electron density for Pks(ACP5+KS6). The $2F_o$-F_c electron density (contoured to 1.0 r.m.s.d.) for the Pks(ACP5+KS6) structure is shown in blue, and the backbone traces of PksKS6 that could be fit to the density are shown in yellow. Despite the extremely high thermal factors for protein atoms in the structure (157.3 $Å^2$), the backbone KS atoms fit the density very well. The eight ACPs in the asymmetric unit could not be located in the remainder of the map. A black arrow indicates the LINKS interaction.

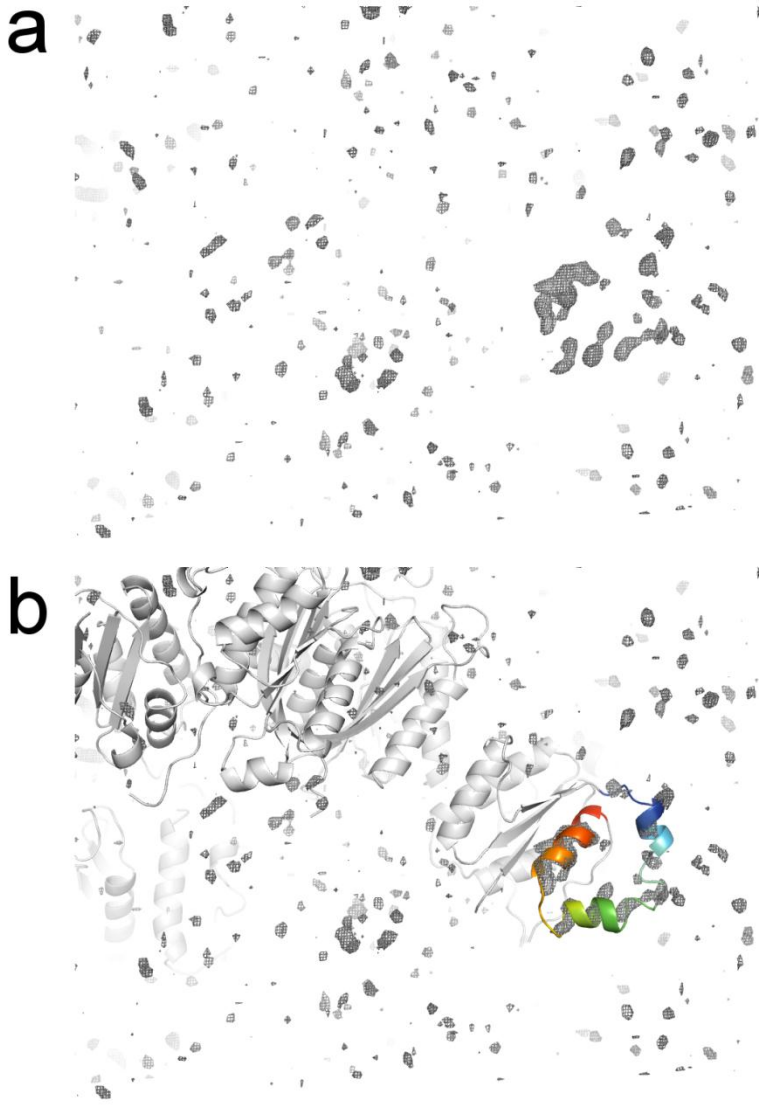
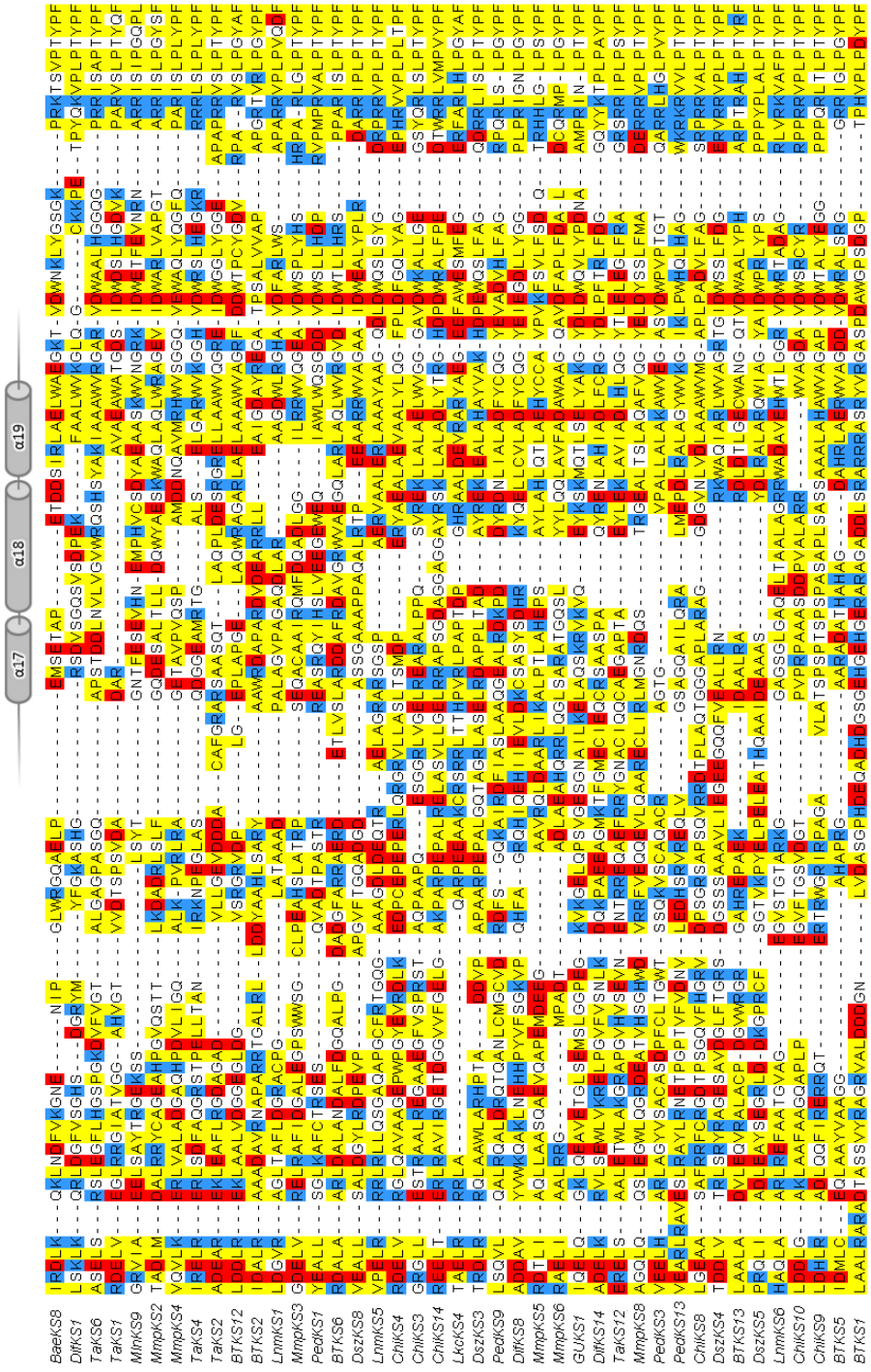
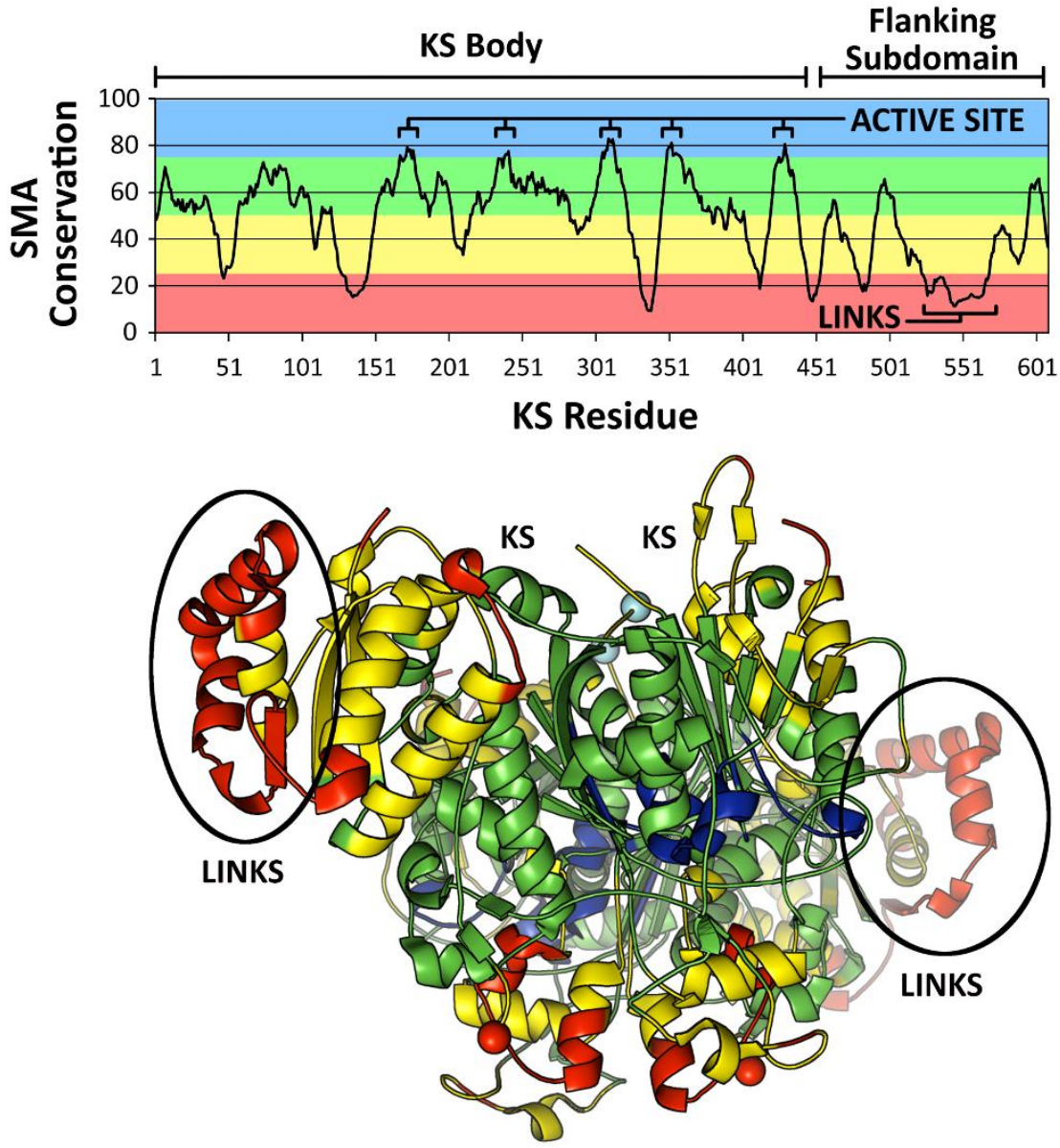


Figure 7-8. F_o-F_c omit map of α17-α19 for BaeKS1. High local B-factors and relatively-poor resolution (3.0 Å) made the construction of α17-α19 for BaeKS1 very challenging. To confirm that the helices have been correctly placed, an F_o-F_c omit map for the region was generated. **(a)** The F_o-F_c omit map contoured at 3.0 r.m.s.d. is shown for a generous portion of the unit cell. The atoms omitted from phasing to produce the map were all those contained in α17-α19. **(b)** To orient the viewer, the model of BaeKS1 has been added into the map, with the deleted helices shown in color.



(Figure 7-9 continued on next page.)

Figure 7-9. Sequence alignment of LINKS region with uncharacterized binding interactions. For approximately half of the sequences used in the alignment of over 100 *trans*-AT KSs, the LINKS region does not reveal the conserved attributes observed in Figure 7-6. It is expected that a structurally-uncharacterized form of LINKS interaction occurs in these KSs, but it is not clear from the sequence alignment how this interface forms. The relative positioning of α17-α19 from Figure 7-6 is shown again for reference.



(Figure 7-10 continued on next page.)

Figure 7-10. Simple moving average calculation of sequence conservation. To determine which regions of *trans*-AT KSs are the most susceptible to sequence variation, a simple moving average (SMA) calculation was performed based on sequence conservation and residues were colored according to the corresponding quartiles. The five regions that received the highest SMA conservation scores all map to the active site of the KS body (blue), while the lowest scoring region corresponds to the LINKS helices (red). A representative KS structure is shown colored accordingly. N- and C-termini are depicted by cyan and red spheres, respectively, and residue numbering is based on the structure of PksKS2.

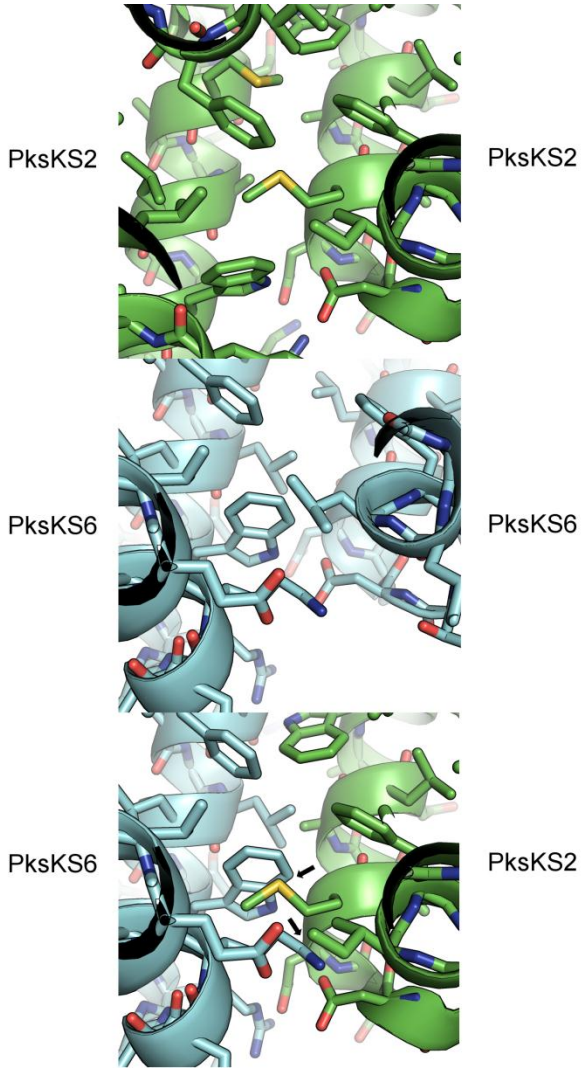
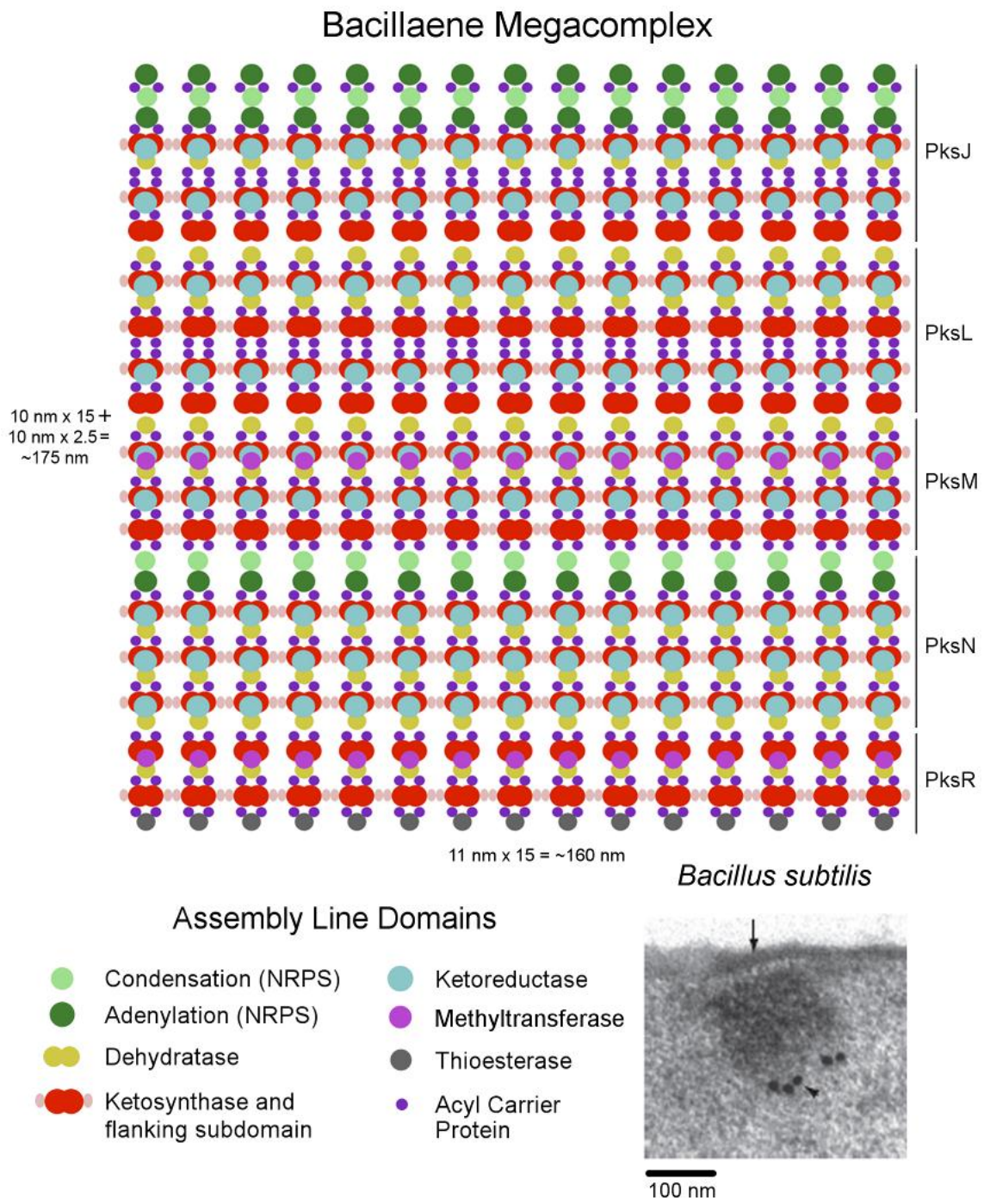


Figure 7-11. Homotypic LINKS interactions are favored over heterotypic. In the top panel, a region of the LINKS interaction for PksKS2 is shown. In the middle panel, the same region of the PksKS6 LINKS interaction is also shown. Modeling a pseudo-LINKS interaction between PksKS6 and PksKS2 reveals several clashes, and two are shown in the bottom panel (black arrows). Both W543 and K551 from PksKS6 are within clashing distance of M559 from PksKS2, responsible for a key hydrophobic contact in the PksKS2 LINKS interaction.



(Figure 7-12 continued on next page.)

Figure 7-12. Vertical and lateral interactions stabilize the PKS megacomplex. The electron micrographs observed by Straight *et al.* revealed a dense mass associated with the membrane of *B. subtilis* cells, identified to consist of numerous copies of the PksX megasynthase (Straight *et al.*, 2007). An image from that work is reprinted here, revealing that the observed superstructure measures approximately 150 × 150 nm. The arrow points to the cell membrane, and the arrowhead points to gold nanoparticles engineered to target PksX proteins. For visualizing how this mass of proteins may self-associate, a diagram shows the assembly line subunits (PksJ, PksL, PksM, PksN, and PksR) interacting vertically through domain-domain interactions, and laterally through the LINKS network. The estimations for megacomplex dimensions that flank the cartoon are calculated based on available crystal structures of individual PKS domains. In the vertical dimension, 175 nm is represented by 15 PKS modules and 2.5 NRPS modules (10 nm each). In the lateral dimension, 160 nm is estimated by the total length of 15 KS homodimers that polymerize through LINKS interactions.

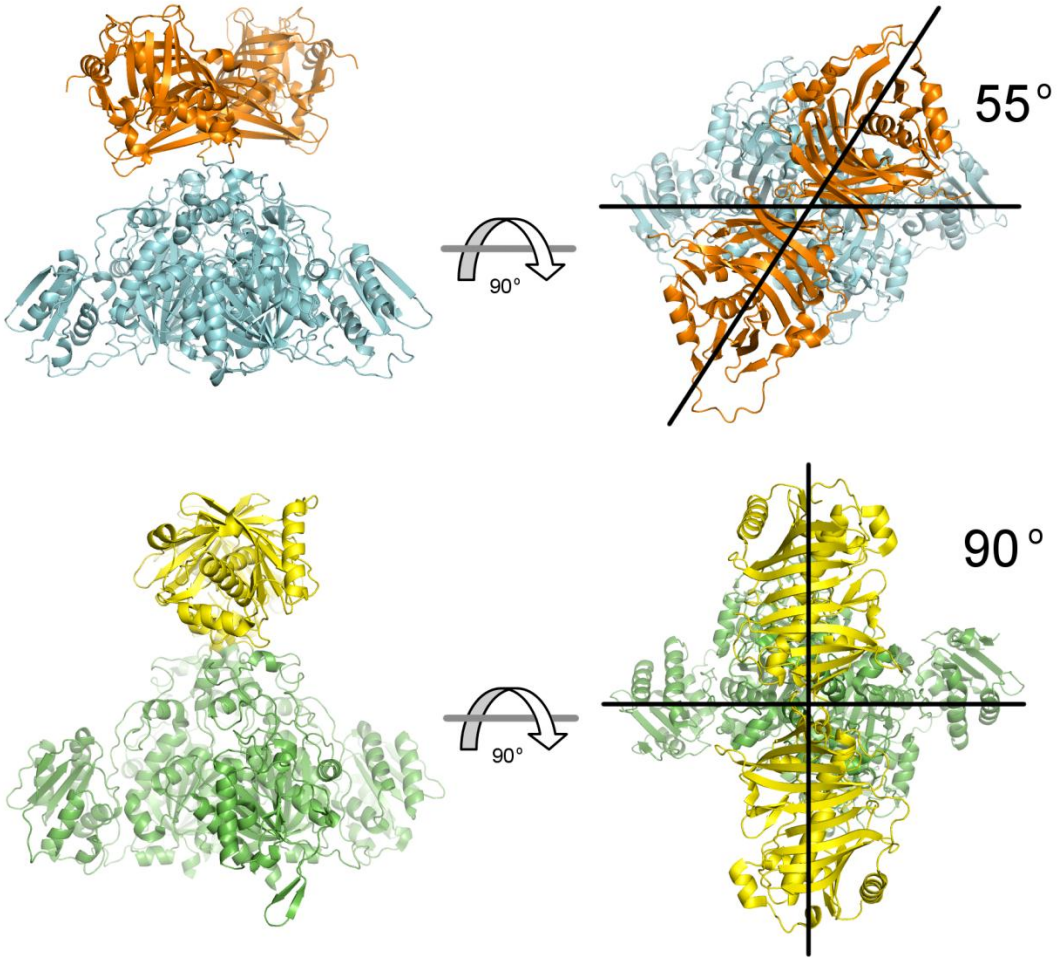
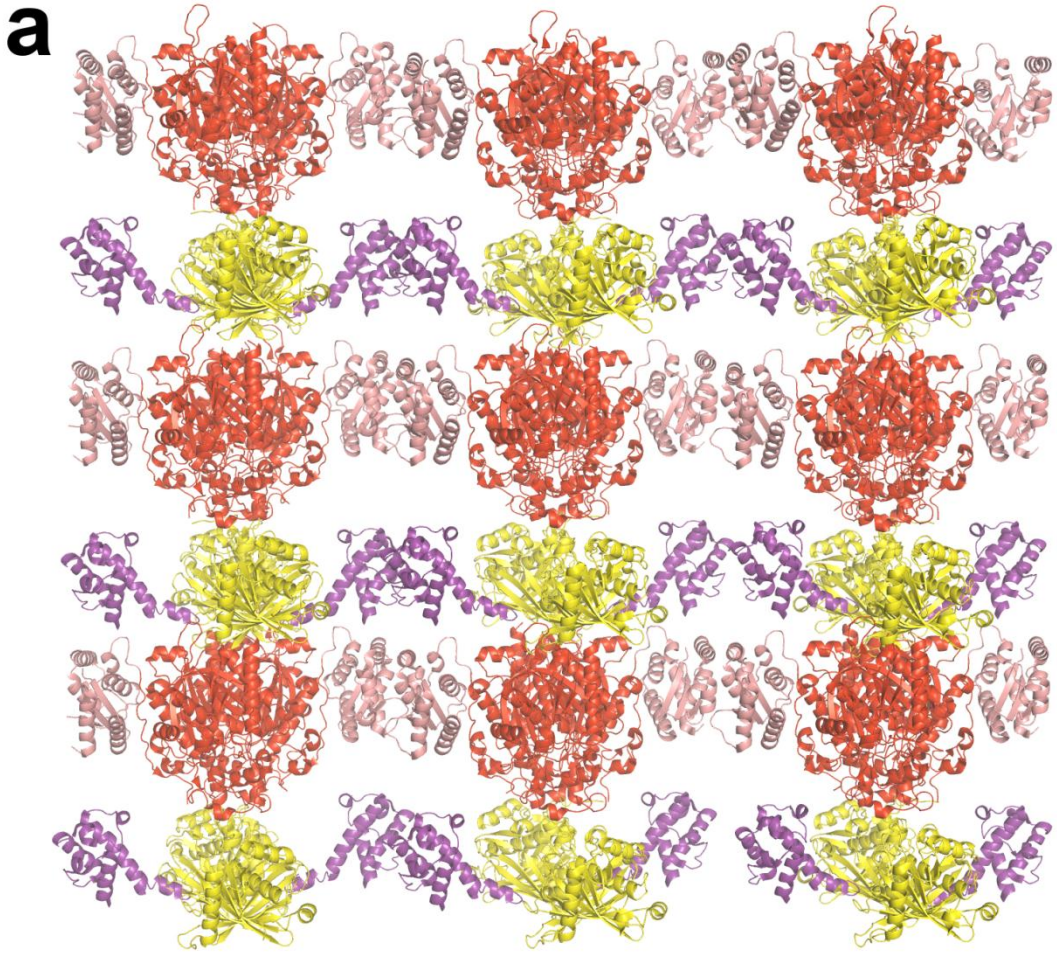
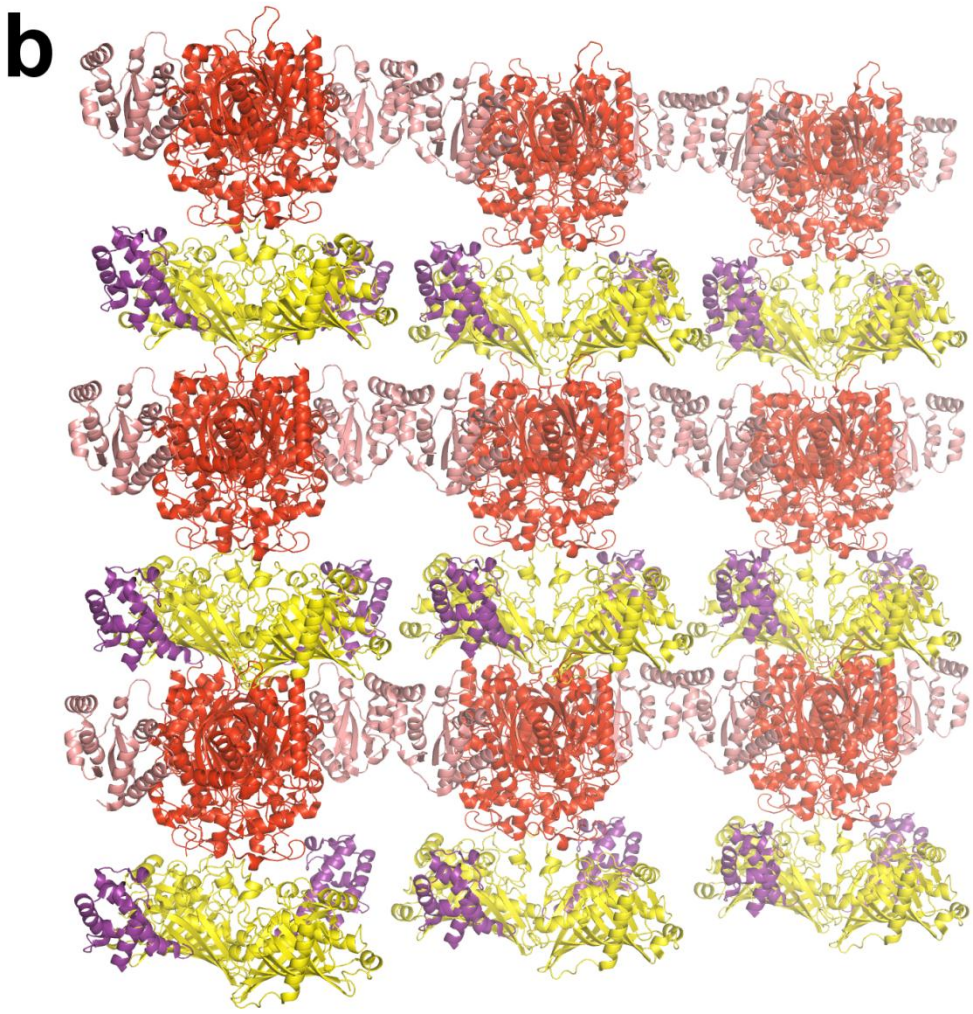


Figure 7-13. Relative orientation of mammalian fatty acid synthase(KS+DH) and Rhi11(KS+B). Shown in cyan and orange are the mammalian fatty acid synthase KS and DH domains, respectively (PDB: 2VZ8). Shown in green and yellow are the *trans*-AT KS and B domains, respectively (PDB: 4KC5). The relative positioning of the DH and B domains respective to their cognate KS domains differs by 35˚. The orientation of the KS+B didomain may indicate the processing domains of *trans*-AT PKSs have adopted this architecture to accommodate lateral LINKS interactions, which would form a network along the horizontal axis of the image.

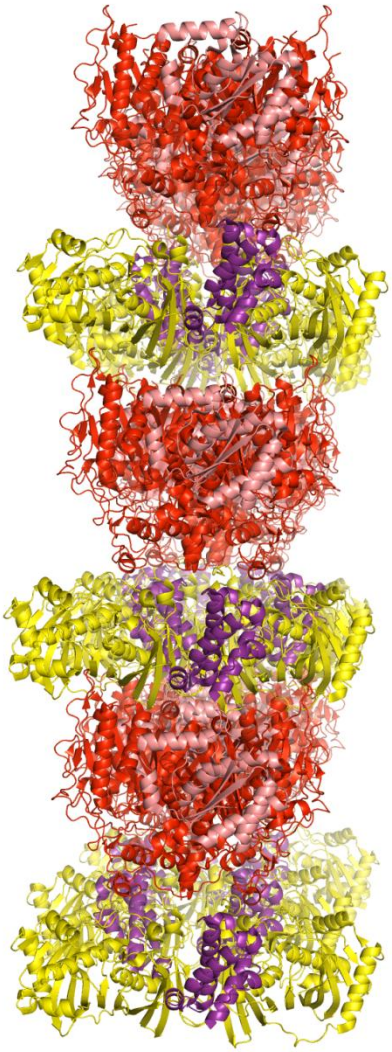


(Figure 7-14 continued on next page.)



(Figure 7-14 continued on next page.)

c



(Figure 7-14 continued on next page.)

Figure 7-14. Protein density of the megacomplex. To reveal the density of the *trans*-AT PKS megacomplex, a simulation of nine modules in megacomplex formation was generated. The modules are based on the crystal structure of Rhi(KS+B)11, PDB: 4KC5 (the B domain is structurally identical to the DH domain). RhiKS11 does not contain LINKS, and was replaced with PksKS2 in each module (red). The flanking subdomains are shown in salmon, the DH domains are shown in yellow, and the ACPs are shown in purple (PDB: 2JU2). **(a)** The nine modules are forming six LINKS interactions. Panels **(b)** and **(c)** represent a 30˚ and 180˚ rotation of the view in panel (a), respectively. Based on the structure of Rhi(KS+B)11, DH domains dimerize orthogonal to KS domains and do not inhibit the LINKS interaction (Figure 7-13). The ACP domains appear to have sufficient freedom to visit the condensation and processing domains of the respective module, as well as the KS of the subsequent module.

	PksKS2 (monomeric)	PksKS6 (Cys169Ser)	PksKS6 (Cys169Ser)	PksKS6	Pks(ACP5+KS6)	BaeKS1	BaeKS5
Data Collection							
Wavelength (Å)	0.9999	1.0332	0.9765	0.9765	1.0332	1.0332	1.0332
Space group	C2	P 2_1	P 2_1	P 1	P 1	P 1	P 2_1
Cell dimensions, *a*, *b*, *c* (Å); α, β, γ (°)	197.1, 75.1, 134.1; 90.0, 132.9, 90.0	89.4, 108.2, 151.2; 90.0, 96.4, 90.0	76.7, 105.5, 89.3; 90.0, 96.8, 90.0	69.0, 77.3, 85.3; 77.7, 70.3, 64.2	63.1, 112.7, 211.4; 105.0, 90.1, 106.3	61.2, 100.0, 100.8; 91.9, 88.2, 96.0	70.8, 319.1, 103.5; 90.0, 110.3, 90.0
Resolution (Å)	47.25-3.02	150.89 -2.00	36.55 - 2.16	39.73 - 3.10	39.77 - 4.00	99.35 - 2.93	97.07 – 4.20
R_{merge}	0.048 (0.663)	0.120 (0.570)	0.092 (0.481)	0.058 (0.359)	0.102 (0.373)	0.148 (0.641)	0.114 (0.483)
$I/\sigma(I)$	32.2 (1.8)	28.1 (2.5)	12.3 (2.0)	13.3 (1.7)	6.1 (2.7)	15.7 (1.5)	14.2 (2.8)
Completeness (%)	99.7 (98.4)	99.2 (96.0)	99.2 (93.5)	94.1 (75.6)	98.7 (99.0)	94.9 (89.75)	79.9 (59.3)
Redundancy	3.7 (3.8)	2.1 (2.1)	1.9 (1.8)	1.9 (1.9)	3.3 (3.5)	1.1 (1.1)	3.2 (2.8)
Wilson B value (Å^2)	119.7	44.4	42.5	77.3	157.2	62.4	192.3
Refinement							
Resolution (Å)	47.25-3.02	150.89 -2.00	36.55 - 2.16	39.73 - 3.10	39.77 - 4.00	99.35 - 2.93	97.07 – 4.20
No. of reflections	26825 (1879)	183336 (13049)	71285 (4947)	24128 (1403)	39874 (2246)	46279 (3209)	23712 (1314)
R_{work}/R_{free}	0.223/0.268	0.201/0.231	0.211/0.248	0.198/0.249	0.318/0.346	0.253/0.283	0.269/0.330
No. of atoms							
Protein	8938	17314	8463	8944	34416	16904	18406
Water	-	1358	88	-	-	-	-
Average *B* factors (Å^2)							
Protein	119.7	44.3	42.7	77.8	157.3	63.1	194.2
Water	-	27	42.4	-	-	-	-
RMS deviations							
Bond lengths (Å)	0.0018	0.0186	0.0083	0.0180	0.0013	0.0127	0.0130
Bond angles (°)	1.084	1.352	1.342	1.390	0.716	1.937	1.818
Ramachandran Statistics (%)							
Preferred Regions	91.8	97.3	96.5	94.0	95.8	92.6	95.8
Allowed Regions	8.2	2.7	3.5	5.9	4.2	6.7	4.2
Outliers	0.0	0.0	0.0	0.1	0.0	0.7	0.0

Table 7-1. Crystallographic data and refinement statistics.

		546	548	552	553	556	557	558	559	562	566	569	571	574	578	581
		N	E	I	F	D	E	D	M	A	W	K	K	K	L	K
548	E	4.0														3.3
552	I				3.5									3.8	4.0	
553	F			4.4					3.7							
556	D										2.9		2.3			
557	E											2.6				
558	D											3.4	4.3			
559	M				4.9				4.5	2.7	2.6					
562	A								3.7							
563	L								4.0							
566	W			4.4		2.9			3.5							
569	K						3.4	2.8								
571	K					2.6		2.8								
574	K					2.8										
578	L			4.7												
581	K															

Hydrogen bond

Hydrophobic int.

Salt bridge

Sulfur-π int.

Table 7-2. Inter-atomic distances for favorable LINKS interactions for PksKS2.

		1106	1107	1108	1110	1113	1116	1117	1120	1122	1123	1126	1130	1138	1142	1145	1147
		R	V	R	T	T	A	L	K	E	L	L	W	D	F	K	M
1106	R					3.3											
1108	R				3.0	2.8											
1110	T			3.0													
1113	T	3.3	4.4	2.8													
1114	L							4.3									
1116	A															3.5	
1117	L							3.7									3.9
1120	K													3.3			
1122	E											4.0	4.0				
1123	L										3.9	3.8			4.2		
1130	W									4.0							
1138	D								3.3								
1142	F										4.2						
1145	K						3.5										
1147	M							3.6									

Hydrogen bond

Hydrophobic int.

Salt bridge

Table 7-3. Inter-atomic distances for favorable LINKS interactions for MgsKS5.

Chapter 8. Conclusions and Future Perspectives

This work examined the structure and function of several enzymes involved in polyketide biosynthesis, and unveiled a motif that forms the structural foundation of the PKS megacomplex. The findings have revealed that harnessing PKS enzymes to generate designer molecules will soon become a reality, as more details regarding their architecture, function, and substrate promiscuity emerge. Whether by reengineering entire synthases to build novel natural products *in vivo* or isolating PKS domains to catalyze valuable reactions *in vitro*, the future of natural product biocatalysis only awaits our own determination and creativity.

In **Chapter 2**, the structure and function of a PKS KR+ER didomain revealed that the PKS and FAS architectures diverged early in evolution, and that analogies may be more difficult to form between the two systems than previously hypothesized. It will not be possible to rationally engineer a man-made PKS without understanding the large-scale ramifications of domain- and module-swapping, and the structure of a complete module would undoubtedly solve many of the outstanding questions. In **Chapter 3**, the stereoselectivity of a PKS DH domain was completely reversed by exchanging the thioester handle attached to the substrate for dehydration. Only the cognate ACP for this DH domain was able to deliver a substrate for dehydration that matches the expected biological selectivity, a finding that reveals potential complications for rational PKS design. In **Chapter 4**, homologous recombination in *S. cerevisiae* was used to generate a new LIC-cloning plasmid, pGAY-28. This plasmid is much more efficient for cloning large PKS genes compared to the more traditional pET-28 series of vectors, demonstrated by the successful incorporation of a 10.3 kb insert. Although the massive size of PKS genes and proteins creates unique obstacles for plasmid generation and protein

purification, the modularity that accompanies the size of these systems is exactly what will be taken advantage of for biocatalytic engineering. In **Chapter 5**, the structure and mechanism of the newly identified enoyl-isomerase domain uncovered a unique mechanism for polyketide double bond migration. The structure revealed a single histidine mediates catalysis unaided by other active site residues or water, shuttling a proton from the γ- position of the enoyl thioester to the α-position. Double bonds control the topography and structural configuration of macrocycles more than any other functional group, and this domain may potentially be employed in biocatalytic reactions to catalyze the migration of double bonds in conjugation with a thioester. In **Chapter 6**, the structure of a PKS KS domain covalently bound to its natural polyketide intermediate answered many questions regarding how polyketides bind in the active site of this domain. The gatekeeping functionality of KS domains may very well be the primary barricade to rational PKS engineering, and experiments using the techniques presented in Chapter 6 to determine how other KS clades select for substrates should be performed. In **Chapter 7**, the LINKS motif of *trans*-AT PKS KS domains was found to mediate the formation of the organelle-sized PKS megacomplex observed by others through cryo-electron microscopy. This motif binds in a homotypic fashion to a spatially-reversed copy of itself, creating a horizontal network that stabilizes the vertical assembly of the PKS megasynthase. While *trans*-AT systems embody an excellent template for the engineering of synthetic PKSs for the exploration of new medicines, it may be necessary to consider the consequences of modifying the LINKS network when relocating modules or domains within a *trans*-AT PKS megasynthase.

References

Adams PD *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **D66**, 213-221.

Agnihotri G & Liu HW (2003) Enoyl-CoA hydratase. Reaction, mechanism, and inhibition. *Bioorg. Med. Chem.* **11**, 9−20.

Airenne TT *et al.* (2003) Structure-function analysis of enoyl thioester reductase involved in mitochondrial maintenance. *J. Mol. Biol.* **327**, 47–59.

Akey DL *et al.* (2010) Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **18**, 94-105.

Alekseyev VY *et al.* (2007) Solution structure and proposed domain-domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci*. **16**, 2093-2107.

Alhamadsheh MM *et al.* (2007) Modular polyketide synthases and *cis* double bond formation: Establishment of activated cis-3-cyclohexylpropenoic acid as the diketide intermediate in phoslactomycin biosynthesis. *J. Am. Chem. Soc.* **129**, 1910−1911.

Anand S & Mohanty D (2012) Modeling holo-ACP:DH and holo-ACP:KR complexes of modular polyketide synthases: A docking and molecular dynamics study. *BMC Struct. Biol.* **12**, 10.

Aron ZD *et al.* (2007) FenF: Servicing the mycosubtilin synthetase assembly line *in trans*. *ChemBioChem* **8**, 613-616.

Aslanidis C & Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069–6074.

August PR *et al.* (1998) Biosynthesis of the ansamycin antibiotic rifamycin: Deductions from the molecular analysis of the *rif* biosynthetic gene cluster of Amycolatopsis mediterranei S699. *Chem. Biol.* **5**, 69−79.

Baerga-Ortiz A *et al.* (2006) Directed mutagenesis alters the stereochemistry of catalysis by isolated ketoreductase domains from the erythromycin polyketide synthase. *Chem. Biol.* **13**, 277–285.

Bahnson BJ & Anderson VE (1991) Crotonase-catalyzed β-elimination is concerted: A double isotope effect study. *Biochemistry* **30**, 5894−5906.

Bevitt DJ *et al.* (1993) Mutagenesis of the dehydratase active site in the erythromycin-producing polyketide synthase. *Biochem. Soc. Trans.* **21**, 30S.

Bonnett SA *et al.* (2013) Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a *cis*-alkene. *Chem. Biol.* **20**, 772−783.

Bradford M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248−254.

Bretschneider T *et al.* (2013) Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **502**, 124−128.

Brignole EJ *et al.* (2009) Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nat. Struct. Mol. Biol.* **16**, 190-197.

Broadhurst RM *et al.* (2003) The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723-731.

Brookes E *et al.* (2009) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. *Eur. Biophys. J.* **39**, 405-14.

Brückner R *et al.* (2008) Praktikum präparative organische chemie. (Spektrum: Heidelberg, 2008) 296.

Butcher RA *et al.* (2007) The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis. *Proc. Natl. Acad. Sci. USA* **104**, 1506–1509.

Caffrey P (2003) Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem* **4**, 654−657.

Calderone CT *et al.* (2006) Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis. *Proc. Natl. Acad. Sci. USA* **103**, 8977–8982.

Calderone CT *et al.* (2008) A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both alpha- and beta-ketone reduction during chain growth. *Proc. Natl. Acad. Sci. USA* **105**, 12809–12814.

Cane DE (2010) Programming of erythromycin biosynthesis by a modular polyketide synthase. *J. Biol. Chem.* **285**, 27517−27523.

Castonguay R *et al.* (2007) Stereospecificity of ketoreductase domains of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **129**, 13758−13769.

Castonguay R *et al.* (2008) Stereospecificity of ketoreductase domains 1 and 2 of the tylactone modular polyketide synthase. *J. Am. Chem. Soc.* **130**, 11598−11599.

Chakravarty B *et al.* (2004) Human fatty acid synthase: structure and substrate selectivity of the thioesterase domain. *Proc. Natl. Acad. Sci. USA* **101**, 15567–15572.

Chen AY *et al.* (2006) Extender unit and acyl carrier protein specificity of ketosynthase domains of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **128**, 3067−3074.

Chen AY *et al.* (2007) Structure-based dissociation of a type I polyketide synthase module. *Chem. Biol.* **14**, 784−792.

Chen XH *et al.* (2006) Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42. *J. Bacteriol.* **188**, 4024–4036.

Chen ZJ *et al.* (2008) Structural enzymological studies of 2-enoyl thioester reductase of the human mitochondrial FAS II pathway: new insights into its substrate recognition properties. *J. Mol. Biol.* **379**, 830–844.

Cheng YQ *et al.* (2003) Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. USA* **100**, 3149–3154.

Chino A *et al.* (2010) Plasmid construction using recombination activity in the fission yeast Schizosaccharomyces pombe. *PLoS ONE* **5**, e9652.

Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr.* **D50**, 760−763.

Davison J *et al.* (2014) Insights into the function of *trans*-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chem. Sci.* **5**, 3081-3095.

De Marco V *et al.* (2004) The solubility and stability of recombinant proteins are increased by their fusion to NusA. *Biochem. Biophys. Res. Commun.* **322**, 766−771.

Demeler B & van Holde KE (2004) Sedimentation velocity analysis of highly heterogeneous systems. *Anal. Biochem.* **335**, 279-288.

Demeler B. (2005) A comprehensive data analysis software package for analytical ultracentrifugation experiments, in *Modern analytical ultracentrifugation: techniques and methods.* (Scott, D.J., Harding, S.E. & Rowe, A.J., Eds.) 210-229 (Royal Society of Chemistry, UK).

Demeler B & Brookes E (2008) Monte Carlo analysis of sedimentation experiments. *Colloid. Polym. Sci.* **286**, 129-37.

Deska J *et al.* (2011) Stereoselective synthesis of deuterated β-cyclohexenylserine, a biosynthetic intermediate of the salinosporamides. *Org. Lett.* **13**, 3210–3213.

Donadio S *et al.* (1993) An erythromycin analog produced by reprogramming of polyketide synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7119−7123.

Dutta S *et al*. (2014) Structure of a modular polyketide synthase. *Nature* 510, 512-517.

Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Cryst.* **D60**, 2126-2132.

Fischer H *et al.* (2010) Determination of the molecular weight of proteins in solution from a single small-angle x-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109.

Floss HG & Yu TW (1999) Lessons from the rifamycin biosynthetic gene cluster. *Curr. Opin. Chem. Biol.* **3**, 592−597.

Franke D & Svergun DI (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346.

Gay D *et al.* (2013) Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. *Biochemistry* **52**, 8916−8928.

Gay DC *et al.* (2014) A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. *Structure* **22**, 444–451.

Gay G *et al.* (2014) Rapid modification of the pET-28 expression vector for ligation-independent cloning using homologous recombination in Saccharomyces cerevisiae. *Plasmid* **76**, 66-71.

Gay DC *et al.* (2014) A double-hotdog with a new trick: structure and mechanism of the *trans*-acyltransferase polyketide synthase enoyl-isomerase. *ACS Chem. Biol.* **9**, 2374-2381.

Gietz RD & Akio S (1989) New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene* **74**, 527–534.

Gomes B *et al.* (1981) Mechanism of action of glutaryl-CoA and butyryl-CoA dehydrogenases. Purification of glutaryl-CoA dehydrogenase. *Biochemistry* **20**, 1481-90.

Guo X *et al.* (2010) Mechanism and stereospecificity of a fully saturating polyketide synthase module: Nanchangmycin synthase module 2 and its dehydratase domain. *J. Am. Chem. Soc.* **132**, 14694−14696.

Guo X *et al.* (2012) Essential role of the donor acyl carrier protein in stereoselective chain translocation to a fully reducing module of the nanchangmycin polyketide synthase. *Biochemistry* **51**, 879−887.

Haapalainen AM *et al.* (2007) Crystallographic and kinetic studies of human mitochondrial acetoacetyl-CoA thiolase: The importance of potassium and chloride ions for its structure and function. *Biochemistry* **46**, 4305−4321.

Hantash FM & Earhart CM (2000) Membrane association of the Escherichia coli enterobactin synthase proteins EntB/G, EntE, and EntF. *J. Bacteriol.* **182**, 1768-1773.

Haun RS *et al.* (1992) Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. *Biotechniques* **13**, 515–518.

He H-Y *et al.* (2014) *Cis* double bond formation by thioesterase and transfer by ketosynthase in FR901464 biosynthesis. *J. Am. Chem. Soc.* **136**, 4488−4491.

Hertweck C (2009) The biosynthetic logic of polyketide diversity. *Angew. Chem., Int. Ed.* **48**, 4688−4716.

Holm L & Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-W549.

Hünig S *et al.* (2006) Arbeitsmethoden in der organischen Chemie. 2 (Verlag Lehmanns: Berlin) 52–272.

Imperi F & Visca P (2013) Subcellular localization of the pyoverdine biogenesis machinery of *Pseudomonas aeruginosa*: A membrane-associated "siderosome". *FEBS Lett.* **587**, 3387-3391.

Jefford C & Wang Y (1987) A simple general distereoselective synthesis of 5-hydroxyalkylbutan-4-olides. *J. Chem. Soc., Chem. Commun.* 1987, 1513–1514.

Jenner M *et al.* (2013) Substrate specificity in ketosynthase domains from *trans*-AT polyketide synthases. *Angew. Chem. Int. Ed. Engl.* **52**, 1143– 1147.

Jez JM *et al.* (2000). Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **39**, 890-902.

Kabsch W & Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.

Kao CM *et al.* (1998) Alcohol stereochemistry in polyketide backbones is controlled by the β-ketoreductase domains of modular polyketide synthases. J. Am. Chem. Soc. **120**, 2478-2479.

Kapur S *et al.* (2012) Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4110−4115.

Kasaragod P *et al.* (2013) The isomerase and hydratase reaction mechanism of the crotonase active site of the multifunctional enzyme (type-1), as deduced from structures of complexes with 3S-hydroxy-acyl-CoA. *FEBS J.* **280**, 3160−3175.

Keatinge-Clay AT *et al.* (2003) Catalysis, specificity, and ACP docking site of Streptomyces coelicolor malonyl-CoA:ACP transacylase. *Structure* **11**, 147–154.

Keatinge-Clay AT & Stroud RM (2006) The Structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. *Structure* **14**, 737-748.

Keatinge-Clay AT (2007) A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **14**, 898-908.

Keatinge-Clay AT (2008) Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384**, 941-953.

Keatinge-Clay AT (2012) The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050−1073.

Khosla C *et al.* (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem.* **76**, 195-221.

Khosla C *et al.* (2014) Assembly line polyketide synthases: mechanistic insights and unsolved problems. *Biochemistry* **53**, 2875-2883.

Kim CY *et al.* (2004) Reconstituting modular activity from separated domains of 6-deoxyerythronolide B synthase. *Biochemistry* **43**, 13892−13898.

Kim KH *et al.* (2011) Crystal structures of enoyl-ACP reductases I (FabI) and III (FabL) from B. subtilis. *J. Mol. Biol.* **406**, 403–415.

Kimber MS *et al.* (2004) The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from Pseudomonas aeruginosa. *J. Biol. Chem.* **279**, 52593−52602.

Kirst H (2010) The spinosyn family of insecticides: realizing the potential of natural products research. *J. Antibiot.* **62**, 101-111.

Kishan KV *et al.* (1997) The SH3 domain of Eps8 exists as a novel intertwined dimer. *Nat. Struct. Biol.* **4**, 739–743.

Kohlhaas C *et al.* (2013) Substrate specificity of ketosynthase domains in *trans*-AT polyketide synthases: amino acid-containing acyl chains. *Chem. Sci.* **4**, 3212–3217.

Konarev PV *et al.* (2003) PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst*. **36**, 1277-82.

Kong R *et al.* (2013) Elucidation of the biosynthetic gene cluster and the post-PKS modification mechanism for fostriecin in Streptomyces pulveraceus. *Chem. Biol.* **20**, 45−54.

Kozin M & Svergun D (2000) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41.

Kumar P *et al.* (2003) Intermodular communication in modular polyketide synthases: Structural and mutational analysis of linker mediated protein-protein recognition. *J. Am. Chem. Soc.* **125**, 4097-4102.

Kusebauch B *et al.* (2010) Functionally distinct modules operate two consecutive α,β → β,γ double bond shifts in the rhizoxin polyketide assembly line. *Angew. Chem.* **122**, 1502−1506.

Kwan DH *et al.* (2008) Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases. *Chem. Biol.* **15**, 1231–1240.

Kwan DH & Leadlay PF (2010) Mutagenesis of a modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. *ACS Chem. Biol.* **5**, 829–838.

Lewy DS *et al.* (2002) Fostriecin: Chemistry and biology. *Curr. Med. Chem.* **9**, 2005−2032.

Liu Y & Eisenberg D (2002) 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**, 1285–1299.

Liu W *et al.* (2007) Helicobacter pylori acyl carrier protein: Expression, purification, and its interaction with β-hydroxyacyl-ACP dehydratase. *Protein Expression Purif.* **52**, 74−81.

Liu X *et al.* (2007) The effect of deuteration on protein structure: a high-resolution comparison of hydrogenous and perdeuterated haloalkane dehalogenase. *Acta Crystallogr.* **D63**, 1000−1008.

Lohman JR *et al.* (2013) Structure of the bifunctional acyltransferase/decarboxylase LnmK from the leinamycin biosynthetic pathway revealing novel activity for a double-hot-dog fold. *Biochemistry* **52**, 902−911.

Lohr F *et al.* (2013) α,β → β,γ double bond migration in corallopyronin A biosynthesis. *Chem. Sci.* **4**, 4175−4180.

Lopanik NB *et al*. (2008) *In vivo* and *in vitro trans*-acylation by BryP, the putative bryostatin pathway acyltransferase derived from an uncultured marine symbiont. *Chem. Biol.* **15**, 1175-1186.

Lu H & Tonge PJ (2010) Mechanism and inhibition of the FabV enoyl-ACP reductase from Burkholderia mallei. *Biochemistry* **49**, 1281–1289.

Luna-Vargas MPA *et al.* (2011) Enabling high-throughput ligation independent cloning and protein expression for the family of ubiquitin specific proteases. *J. Struct. Biol.* **175**, 113–119.

Ma H *et al.* (1987) Plasmid construction by homologous recombination in yeast. *Gene* **58**, 201–216.

Ma SM *et al.* (2009) Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **326**, 589–592.

Maier T *et al.* (2008) The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-1322.

McCoy AJ *et al.* (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658−674.

Menche D *et al.* (2009). Modular total synthesis of archazolid A and B. *J. Org. Chem.* **74**, 7220-7229.

Moldenhauer J *et al.* (2007) Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the *trans*-AT family. *Angew. Chem. Int. Ed. Engl.* **46**, 8195–8197.

Moldenhauer J *et al.* (2010) The final steps of bacillaene biosynthesis in Bacillus amyloliquefaciens FZB42: Direct evidence for β,γ-dehydration by a *trans*-acyltransferase polyketide synthase. *Angew. Chem.* **122**, 1507− 1509.

Moynié L *et al.* (2013) Structural insights into the mechanism and inhibition of the β-hydroxydecanoyl-acyl carrier protein dehydratase from Pseudomonas aeruginosa. *J. Mol. Biol.* **425**, 365−377.

Münster N *et al.* (2012) Differentiation of diastereotopic bromine atoms in SN2 reactions of gem-dibromides. *Chem. Commun.* **48**, 1866–1867.

Murshudov GN *et al.* (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **D53**, 240−255.

Musiol EM *et al.* (2011) Supramolecular templating in kirromycin biosynthesis: the acyltransferase KirCII loads ethylmalonyl-CoA extender onto a specific ACP of the trans-AT PKS. *Chem. Biol.* **18**, 438–444.

Nakanaga K *et al.* (2013) Buruli ulcer and mycolactone-producing mycobacteria. *Jpn. J. Infect. Dis.* **66**, 83-88.

Nguyen T *et al.* (2008) Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–233.

Noack M & Göttlich R (2002) Iodide-catalysed cyclization of unsaturated N chloroamines: a new way to synthesise 3-chloropiperidines. *Eur. J. Org. Chem.* 3171–3178.

Olano C *et al.* (2004) Biosynthesis of the angiogenesis inhibitor borrelidin by Streptomyces parvulus Tu4055: Cluster analysis and assignment of functions. *Chem. Biol.* **11**, 87−97.

Oldenburg KR *et al.* (1997) Recombination mediated PCR-directed plasmid construction *in vivo* in yeast. *Nucleic Acids Res.* **25**, 451–452.

Olsen JG *et al.* (2001) Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. *Structure* **9**, 233–243.

Oppermann U *et al.* (2003) Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem. Biol. Interact.* **143–144**, 247–253.

Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Page PC *et al.* (2004) The efficient synthesis of alkoxy-esters from hydroxy carboxylic acids using dimsyllithium in dimethylsulfoxide followed by alkylation with an alkyl halide. *Synlett* **14**, 2606–2608.

Palaniappan N *et al.* (2003) Enhancement and selective production of phoslactomycin B, a protein phosphatase IIa inhibitor, through identification and engineering of the corresponding biosynthetic gene cluster. *J. Biol. Chem.* **278**, 35552−35557.

Palaniappan N *et al.* (2008) cis-Δ(2,3)-double bond of phoslactomycins is generated by a post-PKS tailoring enzyme. *J. Am. Chem. Soc.* **130**, 12236−12237.

Pappenberger G *et al.* (2010) Structure of the human fatty acid synthase KS-MAT didomain as a framework for inhibitor design. *J. Mol. Biol.* **397**, 508–519.

Partanen ST *et al.* (2004) The 1.3 Å crystal structure of human mitochondrial Δ^3-Δ^2-enoyl-CoA isomerase shows a novel mode of binding for the fatty acyl group. *J. Mol. Biol.* **342**, 1197−1208.

Pasta S *et al.* (2007) Catalytic residues are shared between two pseudosubunits of the dehydratase domain of the animal fatty acid synthase. *Chem. Biol.* **14**, 1377−1385.

Patel PS *et al.* (1995) Bacillaene, a novel inhibitor of prokaryotic protein synthesis produced by Bacillus subtilis: production, taxonomy, isolation, physico-chemical characterization and biological activity. *J. Antibiot.* **48**, 997−1003.

Pedelacq JD *et al.* (2011) Experimental mapping of soluble protein domains using a hierarchical approach. *Nucleic Acids Res.* **39**, e125.

Persson B *et al.* (2008) Medium- and short-chain dehydrogenase/reductase gene and protein families: the MDR superfamily. *Cell. Mol. Life Sci.* **65**, 3879–3894.

Piasecki SK *et al.* (2011) Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks. *Chem. Biol.* **18**, 1331-1340.

Piasecki SK *et al.* (2014) Structural and functional studies of a *trans*-acyltransferase polyketide assembly line enzyme that catalyzes stereoselective α- and β-ketoreduction. *Proteins: Struct., Funct., Bioinf.,* DOI: 10.1002/prot.24561.

Pidugu LS *et al.* (2009) Analysis of proteins with the "hot dog" fold: Prediction of function and identification of catalytic residues of hypothetical proteins. *BMC Struct. Biol.* **9**, 37.

Piel J (2010) Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996−1047.

Pihko PM *et al.* (2009) Oxyanion holes and their mimics, in Hydrogen Bonding in Organic Synthesis (Pihko, P., Ed.) pp 43−71, Wiley-VCH Verlag GmbH, Weinheim, Germany.

Plaut B & Knowles J (1972) pH-dependence of the triose phosphate isomerase reaction. *Biochem. J.* **129**, 311−320.

Pollack R (2004) Enzymatic mechanisms for catalysis of enolization: ketosteroid isomerase. *Bioorg. Chem.* **32**, 341−353.

Porter JL *et al*. (2013) The cell wall-associated mycolactone polyketide synthases are necessary but not sufficient for mycolactone biosynthesis. *PLoS ONE* **8(7)**: e70520.

Potterton E *et al.* (2003) A graphical user interface to the CCP4 program suite. *Acta Cryst.* **D59**, 1131-1137.

Pulsawat N *et al.* (2007) Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene* **393**, 31-42.

Putnam CD *et al.* (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285.

Reddick JJ *et al.* (2007) PksS from Bacillus subtilis is a cytochrome P450 involved in bacillaene metabolism. *Biochem. Biophys. Res. Commun.* **358**, 363–367.

Reid R *et al.* (2003) A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**, 72−79.

Reynolds KA *et al.* (1993) Mechanistic studies of a Δ^1,Δ^2-cyclohexenylcarbonyl-CoA isomerase catalyzing the penultimate step in the biosynthesis of the cyclohexanecarboxylic acid moiety of ansatrienin A. *J. Nat. Prod.* **56**, 825−829.

Rosenberg AH *et al.* (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**, 125–135.

Sambrook J *et al.* (1989) Molecular Cloning, A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.

Sasso S *et al.* (2012) Microalgae in the postgenomic era: a blooming reservoir for new natural products. *FEMS Microbiol. Rev.* **36**, 761−785.

Schupp T *et al.* (1998) Cloning and sequence analysis of the putative rifamycin polyketide synthase gene cluster from Amycolatopsis mediterranei. *FEMS Microbiol. Lett.* **159**, 201−207.

Sedgwick B *et al.* (1978) Stereochemical course of dehydration catalyzed by the yeast fatty acid synthetase. *J. Chem. Soc., Chem. Commun.* **5**, 193−194.

Shen B (2003) Polyketide biosynthesis beyond the type I, II, and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285-295.

Sherman DH & Smith, JL (2006) Clearing the skies over modular polyketide synthases. *ACS Chem. Biol.* **1**, 505−509.

Simunovic V *et al.* (2006) Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and *trans*-acting acyltransferases. *ChemBioChem* **7**, 1206-1220.

Siskos AP *et al.* (2005) Molecular basis of Celmer's rules: stereochemistry of catalysis by isolated ketoreductase domains from modular polyketide synthases. *Chem. Biol.* **12**, 1145–1153.

Smith AB *et al.* (1997) Total syntheses of (+)-acutiphycin and (+)-trans-20,21 didehydroacutiphycin. *J. Am. Chem. Soc.* **7863**, 10935–10946.

Smith S & Tsai SC (2007) The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041-1072.

Staunton J *et al.* (1996) Evidence for a double-helical structure for modular polyketide synthases. *Nat. Struct. Biol.* **3**, 188–192.

Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst.* **41**, 641-643.

Stols L *et al.* (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif.* **25**, 8–15.

Straight PD *et al.* (2007) A singular enzymatic megacomplex from *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 305-310.

Stratmann A *et al.* (1999) Intermediates of rifamycin polyketide synthase produced by an Amycolatopsis mediterranei mutant with inactivated *rifF* gene. *Microbiology* **145**, 3365−3375.

Studier FW & Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130.

Studier FW *et al.* (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol.* **185**, 60–89.

Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495-503.

Svergun DI *et al.* (1995) CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.

Taft F *et al.* (2009) Timing of the Δ10,12-Δ11,13 double bond migration during ansamitocin biosynthesis in Actinosynnema pretiosum. *J. Am. Chem. Soc.* **131**, 3812−3813.

Tang G-L *et al.* (2004) Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. *Chem. Biol*. **11**, 33-45.

Tang L *et al.* (1998) Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by Amycolatopsis mediterranei. *Gene* **216**, 255−265.

Tang L *et al.* (2000) Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640−642.

Tang L *et al.* (2004) Elucidating the mechanism of cis double bond formation in epothilone biosynthesis. *J. Am. Chem. Soc.* **126**, 46−47.

Tang Y *et al.* (2006) The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci.* **103**, 11124-11129.

Tang Y *et al.* (2007) Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14**, 931–943.

Terwilliger TC *et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Cryst.* **D64**, 61-69.

Terwilliger TC *et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Cryst.* **D65**, 582-601.

Thorn JM *et al.* (1995) Crystal structure of Escherichia coli QOR quinone oxidoreductase complexed with NADPH. *J. Mol. Biol.* **249**, 785–799.

Torkko JM *et al.* (2003) Candida tropicalis expresses two mitochondrial 2-enoyl thioester reductases that are able to form both homodimers and heterodimers. *J. Biol. Chem.* **278**, 41213–41220.

Tran L *et al.,* (2008) linkage mediates communication between ACP and TE domains in modular polyketide synthases. *Chembiochem* **9**, 905-15.

Tsai SC *et al.* (2001) Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc. Natl. Acad. Sci.* **98**, 14808-14813.

Tsuji SY *et al.* (2001) Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* **40**, 2326-2331.

Valenzano CR *et al.* (2009) The biochemical basis for stereochemical control in polyketide biosynthesis. *J. Am. Chem. Soc.* **131**, 18501–18511.

Valenzano CR *et al.* (2010) Stereospecificity of the dehydratase domain of the erythromycin polyketide synthase. *J. Am. Chem. Soc.* **132**, 14697−14699.

Vergnolle O *et al.* (2011) Stereoselectivity of isolated dehydratase domains of the borrelidin polyketide synthase: Implications for cis double bond formation. *ChemBioChem* **12**, 1011−1014.

Volkov VV & Svergun DI (2003) Uniqueness of *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864.

von Wettstein-Knowles P *et al.* (2006) Fatty acid synthesis. Role of active site histidines and lysine in Cys-His-His-type beta-ketoacyl-acyl carrier protein synthases. *FEBS J.* **273**, 695–710.

Waldron C *et al.* (2001) Cloning and analysis of the spinosad biosynthetic gene cluster of Saccharopolyspora spinosa. *Chem. Biol.* **8**, 487-99.

Walsh CT (2008) The chemical diversity of natural-product assembly lines. *Acc. Chem. Res.* **41**, 4-10.

Weinreb PH *et al.* (1998) Stoichiometry and specificity of *in vitro* phosphopantetheinylation and aminoacylation of the valine-activating module of surfactin synthetase. *Biochemistry* **37**, 1575−1584.

World Health Organization (2010) *Standard Treatment Regimens. In Treatment of Tuberculosis Guidelines*, 4th ed., pp 29−30, World Health Organization Press, Geneva.

Winn MD *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta. Crystallogr. D.* **67**, 235-242.

Witkowski A *et al.* (2004) Characterization of the β-carbon processing reactions of the mammalian cytosolic fatty acid synthase: role of the central core. *Biochemistry* **43**, 10458–10466.

Wong FT *et al.* (2011) Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase. *Biochemistry* **50**, 6539–6548.

Woodward RB *et al.* (1981) Asymmetric total synthesis of erythromycin. 1. Synthesis of an erythronolide A secoacid derivative via asymmetric induction. *J. Am. Chem. Soc.* **103**, 3210-3213.

Wu N *et al.* (2001) Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J. Am. Chem. Soc.* **123**, 6465− 6474.

Xiang H, *et al.* (1999) Interchange of catalytic activity within the 2-enoyl-coenzyme A hydratase/isomerase superfamily based on a common active site template. *Biochemistry* **38**, 7638−7652.

Xu X-P *et al*. (2008) Localization of the ActIII actinorhodin polyketide ketoreductase to the cell wall. *FEMS Microbiol. Lett.* **287**, 15-21.

Ye Z *et al.* (2014) Reprogramming acyl carrier protein interactions of an acyl-CoA promiscuous *trans*-acyltransferase. *Chem. Biol.* **21**, 636-646.

You YO *et al.* (2013) Stereochemistry of reductions catalyzed by methyl-epimerizing ketoreductase domains of polyketide synthases. *J. Am. Chem. Soc.* **135**, 7406−7409.

Yu T-W *et al.* (1999) Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9051−9056.

Zhang D *et al.* (2001) $\Delta^{3,5}$,$\Delta^{2,4}$-dienoyl-CoA Isomerase is a multifunctional isomerase: a structural and mechanistic study. *J. Biol. Chem.* **276**, 13622−13627.

Zhang Y-M *et al.* (2006) Roles of the active site water, histidine 303, and phenylalanine 396 in the catalytic mechanism of the elongation condensing enzyme of Streptococcus pneumoniae. *J. Biol. Chem.* **281**, 17390–17399.

Zheng J *et al.* (2010) Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase. *Structure* **18**, 913-922.

Zheng J & Keatinge-Clay AT (2011) Structural and functional analysis of C2-type ketoreductases from modular polyketide synthases. *J. Mol. Biol.* **410**, 105–117.

Zheng J *et al.* (2012) Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615-621.

Zhu Y & Burgess K (2008) Iridium-catalyzed asymmetric hydrogenation of vinyl ethers. *Adv. Synth. Catal.* **350**, 979–983.